

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: April 6, 2004, 17:52:00 ; Search time 43.5 Seconds  
(without alignments)  
7050.200 Million cell updates/sec

Title: US-10-057-510-1

Perfect score: 876

Sequence: 1 gcacgaggtatnaaaat.....aaagggaantccccatggg 486

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 2034082

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp  
-Q/cgn2.1/USPTC.spool.p/US10057510/runat\_06042004\_141608\_14722/app\_query.fasta\_1.647  
-DB=SPTREMBL\_25 -QWMT=Eastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFWT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10057510@cgn 1.1.86 @runat\_06042004\_141608\_14722 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPTREMBL\_25:

1: sp\_archaea:  
2: sp\_bacteria:  
3: sp\_fungi:  
4: sp\_human:  
5: sp\_invertebrate:  
6: sp\_mammal:  
7: sp\_mhc:  
8: sp\_organalle:  
9: sp\_phage:  
10: sp\_plan:  
11: sp\_rodent:  
12: sp\_virus:  
13: sp\_vertebrate:  
14: sp\_unclassified:  
15: sp\_virus:  
16: sp\_bacteriap:  
17: sp\_archaeap:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	100.5	12.0	132	11 Q8BRC3	Q8BRC3 mus musculus

C 2	98	11.7	258	12	Q917X0
C 3	94.5	11.2	633	10	Q7XI88
C 4	92	11.0	191	4	Q99491
C 5	92	11.0	258	12	Q917W7
C 6	91	10.8	495	5	Q8MSC6
C 7	91	10.8	1355	5	Q9VKH2
C 8	91	10.8	3469	5	Q9U4I2
C 9	90.5	10.8	209	12	Q69115
C 10	90.5	10.8	209	12	Q69123
C 11	89.5	10.7	647	16	Q8XZN8
C 12	89.5	10.7	2414	4	Q9HCL7
C 13	88.5	10.5	1093	5	Q9V6T2
C 14	88.5	10.5	1150	3	Q872W5
C 15	86	10.2	378	16	Q8ZL04
C 16	86	9.8	805	13	Q9PTY3
C 17	85.5	10.2	1325	5	Q9BKV7
C 18	85.5	10.2	1405	2	Q8KMK1
C 19	84.5	10.1	439	10	Q7XEA2
C 20	84.5	10.1	1173	11	Q63624
C 21	84.5	10.1	1427	2	Q8GIZ2
C 22	84.5	10.1	2957	5	Q61845
C 23	84	9.6	336	4	Q7Z3T3
C 24	84	10.0	381	11	Q8C096
C 25	84	10.0	429	13	Q7ZWV6
C 26	84	9.6	696	16	Q7VY84
C 27	83.5	9.9	274	10	Q9M4M5
C 28	83.5	9.5	336	4	Q81WP5
C 29	83.5	9.9	363	5	Q9Y075
C 30	83.5	9.9	817	3	Q07229
C 31	83.5	9.9	3604	5	Q9VYK0
C 32	83.5	9.9	230	4	Q96D64
C 33	83	9.9	470	4	Q81YD6
C 34	83	9.9	1706	16	Q9RZJ2
C 35	83	9.9	1791	5	Q9U6D4
C 36	82.5	9.8	325	10	Q9ZV88
C 37	82.5	9.8	545	3	Q9HEK1
C 38	82.5	9.8	749	10	C04387
C 39	82.5	9.8	873	5	Q9Y076
C 40	82.5	9.8	283	2	O31397
C 41	82	9.4	479	16	Q9RKZ0
C 42	82	9.4	811	13	Q7ZWS6
C 43	82	9.8	414	10	Q94AS1
C 44	81.5	9.7	478	12	Q91TL3
C 45	81.5	9.3			

#### ALIGNMENTS

RESULT 1

ID	Q8BRC3	PRELIMINARY;	PRT;	132 AA.
AC	Q8BRC3;			
DT	01-MAR-2003 (TrEMBLrel. 23, Created)			
DT	01-MAR-2003 (TrEMBLrel. 23, Last sequence update)			
DT	01-MAR-2003 (TrEMBLrel. 23, Last annotation update)			
DE	Hypothetical protein.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
EN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J;			
RX	MEDLINE=22354693; PubMed=12466851;			
RA	The FANTOM Consortium,			
RA	the RIKEN Genome Exploration Research Group Phase I & II Team;			
RT	"Analysis of the mouse transcriptome based on functional annotation of			
RT	60,770 full-length cDNAs."			
RL	Nature 420:563-573(2002).			
DR	EMBL; AK045144; BAC32238.1; -			
KW	Hypothetical protein.			
SQ	SEQUENCE 132 AA; 13680 MW; FD20A00CA4B554C CRC64;			

Alignment Scores: 0.0548 Length: 132  
Pred. No.: 100.50 Matches: 43  
Score: 33.33% Conservative: 6  
Percent Similarity: 29.25% Mismatches: 56  
Best Local Similarity: 11.96% Indels: 42  
Query Match: 11 Gaps: 6  
DB: 11

US-10-057-510-1 (1-486) x Q8BRC3 (1-132)  
QY 473 TCCTTTCTTCAGGCTTGCNAAAAACCGGGGNCNTANGNCCAGTTTTCGAGAACAGA 414  
Db 10 SerLeuArgSerGlyValGlyLeuSerProSerHisSerArgProArgValSerProArg 29  
QY 413 GGGCCAGGACACACCA-----GGGAGGGGGCTTCTTATGCAAGTTCCTCCAT 363  
Db 30 GlyProGlyProAlaProLysSerGlyArg----- 40  
QY 362 GCCATATGAGCCAAAGNCTNGTGGTCGMAACCCCTTGACTGGGCAATGTCTCAGAGTCTG 303  
Db 41 -----ProGlyAlaLeuProAlaSerLeu 48  
QY 302 GGGACNATCCAGGGGTGCATCTTCATCTCAGCCCTGGTA----- 261  
Db 49 AlaValProPro-----ProProSerCysLeuTrpProGluArgGlnGln 63  
QY 260 GCCTTCTGCCCTCCTCCAGGGCTCGTCCATTCTTCAGCTCCAGCGCTCCTCCACAG 201  
Db 64 ArgLeuAlaProArgProSerThrArgThrPheLeuAsnValProSerArgAla---Ala 82  
QY 200 TCCATGCTCCACCTTGCACATGGANACAGCGGTGCCCCACACAGCAGGGAACAA 141  
Db 83 SerLeuProAspLeuCysSerGlyThrSerSerAlaProLeuSer-ArgGlyLysAl 102  
QY 140 CGAAGCT-----GGCAAGAAAGAACAGCAGTATAGAGCGGGGCAACAATG 96  
Db 102 aArgAlaAspGlyArgGlySerGlyProProGluProGlnMetValSerMetTrpGlnTr 122  
QY 95 GAGAAAGAGTATCTATTA 77  
Db 122 pArgGlyGlnAlaAlaLeu 128

RESULT 2  
QY17X0 ID Q917X0 PRELIMINARY; PRT; 258 AA.  
AC Q917X0;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE Minor structural protein.  
GN ORF3.  
OS Norwalk-like virus NLV/New Orleans/306/1994/US.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Caliciviridae;  
OC Norcovirus.  
OX NCBI\_TaxID=171854;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN-Hu/NLV/New Orleans/306/1994/US;  
RX MEDLINE=97193806; PubMed=9041391;  
RA Ando T., Monroe S.S., Noel J.S., Glass R.I.;  
RT "A one-tube method of reverse transcription-PCR to efficiently amplify  
a 3-kilobase region from the RNA polymerase gene to the poly(A) tail  
of small round-structured viruses (Norwalk-like viruses).";  
RN J. Clin. Microbiol. 35:570-577(1997).  
RN (2)  
RP SEQUENCE FROM N.A.  
RC STRAIN-Hu/NLV/New Orleans/306/1994/US;  
RX MEDLINE=20266071; PubMed=10804147;  
RA Ando T., Noel J.S., Fankhauser R.L.;  
RT "Genetic classification of 'Norwalk-like viruses.'";  
RN J. Infect. Dis. 181:S336-S348(2000).  
RN (3)  
RP SEQUENCE FROM N.A.

RC STRAIN-Hu/NLV/New Orleans/306/1994/US;  
RA Ando T., Seto Y., Noel J.S., Monroe S.S., Glass R.I., Fankhauser R.L.;  
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF414422; AAL13020.1; -  
DR InterPro; IPR004278; RNA\_capsid.  
DR Pfam; PF03035; RNA\_capsid; 1.  
SQ SEQUENCE 258 AA; 27412 MW; 272E2A8FA8EA2A4 CRC64;  
Alignment Scores: 0.112 Length: 258  
Pred. No.: 98.00 Matches: 41  
Score: 40.49% Conservative: 25  
Percent Similarity: 25.15% Mismatches: 64  
Query Match: 11.67% Indels: 33  
DB: 12 Gaps: 4

US-10-057-510-1 (1-486) x Q917X0 (1-258)  
QY 447 CAGGGGNCNTANGNCCAGTTTTCAGAACAGAGGGCCAGCACCAACCCAGGAGGGG 388  
Db 108 ProGlySerMet-----LysThrThrSerTySerGly 118  
QY 387 GCTTCTTATGCAAGTTCCTTCATGCTATGATGAGCCAGNCCTNGTGGCTCGAACCCCT 328  
Db 119 LysPheValSerMetAsnPro-----ValArgGlnVal 129  
QY 327 GACTGGGCAATGTCTGAGAGTCTGGGAGCAGATCCAGGGGTGCATCTTCATCTCCAGCC 268  
Db 130 GluPheProGlnProLysSerAlaAlaLeuProSerSerAlaSerValSerSer 149  
QY 267 CTGCTAGCTTCTGCCCTCCTCCAGGGCTCGTCCATTCTTCAGCTCCAGCGCTCTCT 208  
Db 150 GlyArgThrAsnLeuThrAsnSerThrGlnSerThrValSerThrSerSerAlaPro 169  
QY 207 CCACAGC-----TCCATGCTTCCACCTTGCACATCGANACAGCCCTGCCCCAC 157  
Db 170 Pro-SerArgGlySerSerAlaProSerThrLeuSerArgAlaThrThrArgThrSerAs 189  
QY 156 AGCCAGAGGGAACAAACGAAGCTGCGAAGAAAGA----- 123  
Db 189 nTrpValGluGlnAsnArgAsnLeuGluProTyMetArgGlyAlaLeuGlnThrTh 209  
QY 122 -----ACAGCAGTTAGAGCGGGCAACAATGGAGAAAGAGATATCTATATGAGCTT 70  
Db 209 rPheValThrProProSerSerArgAlaSerSerAsnGlyThrValSerThrValProLy 229  
QY 69 GGCATTTTACACAGCAGCTTCCAGCTTTTCCATAAACATGGAACATTTTATATA 10  
Db 229 sGlyValLeuAspSerTrpThrProAlaPheAsnThrArgGlnProLeuPheAlaTy 249  
QY 9 CGCTCGT 3  
Db 249 rLeuArg 251

RESULT 3  
QY17X0 ID Q7X188 PRELIMINARY; PRT; 633 AA.  
AC Q7X188;  
DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Zinc finger protein-like protein.  
GN P045303.32.  
OS Oryza sativa (japonica cultivar-group).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzaceae; Oryza.  
OX NCBI\_TaxID=39947;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Nipponbare;  
RA Sasaki T., Matsumoto T., Yamamoto K.;  
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 7, PAC

RT clone: P0453G03.1;  
 RL Submitted (Oct-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF004276; BAC79830.1; -- EDB82B36082A61F0 CRC64;  
 SQ SEQUENCE 633 AA; 67636 MW; EDB82B36082A61F0 CRC64;

## Alignment Scores:

Pred. No.: 0.303 Length: 633  
 Score: 94.50 Matches: 36  
 Percent Similarity: 37.69% Conservative: 13  
 Best Local Similarity: 27.69% Mismatches: 40  
 Query Match: 11.25% Indels: 41  
 DB: 10 Gaps: 6

US-10-057-510-1 (1-486) x Q7X188 (1-633)

QY 483 ATGGGGANTTTCCCTTTCTTCCAGGCTTTCNAAAACACAGGGGNGCTANGANCACAGTTT 424  
 DB 398 IIEGlyValAspProThrPhe---LeuCysLeuAlaIleGlyAlaProSerSerLeuPhe 416  
 QY 423 CCAGAACAGAGAGGCCAGGCCACCAACAGAGGGGGGCTTCTTATGCAAGTTCCCTTCA 364  
 DB 417 ProGln-----ThraAlaSerAspProCysSerPheAlaProProAla 432  
 QY 363 TGCCATATGAGCCAGNCCNTGTGCTCGAACCTTGACTGGGCAATGCTCGAGAGTCT 304  
 DB 433 ProHisMetSerAlaThr-----AlaLeuLeuGlnLysAla 444  
 QY 303 GGGGACNATCCAGGGGCTGCTATCTTCATCTCAGCCCTGCGAGGCTT---CTGCCCTCC 247  
 DB 445 AlaGluValGlyAlaSerGlnSerSerSerPheLeuLysGluPheGlyLeuAlaAla 464  
 QY 246 TCCAGGGGCTGCTCCATCTCTTCAGCT----- 220  
 DB 465 SerThrSerSerSerProSerLysLeuSerGlnGlyArgPheThrThrGlyAsnThr 484  
 QY 219 -----CCAGGGCTCTCCACAGCTCCCTCCACCTTCCACCTTCCACAGG 175  
 DB 485 ProThrThrSerHisProHisProHisProHisProHis----- 497  
 QY 174 ANACAGCGTGGCCCAAGCAGGAGGA 145  
 DB 498 -----ProHisProGlnGly 503

## RESULT 4

Q99491 PRELIMINARY; PRT; 191 AA.  
 AC Q99491;  
 DT 01-MAY-1997 (TREMBLrel. 03, Created)  
 DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)  
 DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)  
 DE DAN15 protein (Fragment).  
 GN DAN15.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97051922; PubMed=8896557;  
 RA Imbert G., Saudou F., Veyt G., Devys D., Trottier Y., Garnier J.M.,  
 RA Weber C., Mandel J.L., Cance G., Abbas N., Duerr A., Didierjean O.,  
 RA Stevanin G., Agid Y., Brice A.;  
 RT "Cloning of the gene for spinocerebellar ataxia 2 reveals a locus with  
 RT high sensitivity to expanded CAG/glutamine repeats.";  
 RL Nat. Genet. 14:285-291(1996).  
 DR EMBL; Y08266; CAA69592.1; --  
 FT NON\_TER 1 1  
 FT NON\_TER 191 191  
 SQ SEQUENCE 191 AA; 20196 MW; E0A628C019583310 CRC64;

## Alignment Scores:

Pred. No.: 0.473 Length: 191  
 Score: 92.00 Matches: 38

Percent Similarity: 41.18% Conservative: 18  
 Best Local Similarity: 27.94% Mismatches: 44  
 Query Match: 10.95% Indels: 36  
 DB: 4 Gaps: 5

US-10-057-510-1 (1-486) x Q99491 (1-191)

QY 450 AAACACAGGGGNGCTANGANCAGTTTTCAGAACAGAGGGCCAGGCCAACACAGGAG 391  
 DB 9 LysSerGlyGly-----SerGluAlaAlaLeuLysGlu 19  
 QY 390 GGGGCTTCTTATGCAAGTCCCTTTCATGCCATATGAGCCAGNCCNTGTGCTCGAAC 331  
 DB 20 GlyLysAlaAlaAlaLeuSerSerSerSerSerSerSerSerSerSerSerSerSer 37  
 QY 330 CTTCAGTGGCAATGCTCGAGAGTCTGCGGAGCNATCCAGGGGCTGCTATCTTCATCTCA 271  
 DB 38 -----AlaAlaSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 53  
 QY 270 GCCCTGGTAGCTTCTGCTCCCTCTCCAGGGCTGCTCCATCTCTTCAGTCCAGGCT 211  
 DB 54 -----GlyLeuLeuProAsnHisLysLeuLysThrValGlyGluAlaProAlaAla 70  
 QY 210 CCTCCACACGCTC-----CATGCCCTTC 190  
 DB 71 ProProHisGlnGlnHisHisHisHisHisHisHisHisHisHisHisHisHisHis 90  
 QY 189 CACCTTCCACATGGANACAGCGCTGCCCCCAACAGCCAGGAGGAAACAGAGCTGGCA 130  
 DB 91 HisLeu--HisHisHisHisHisHisHisHisHisHisHisHisHisHisHisHisHis 107  
 QY 129 AGAAGAACAGCAGTGTAGCAGGGGCAACAATGGAGAAAGGAG 86  
 DB 107 IingGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 121

## RESULT 5

Q917W7 PRELIMINARY; PRT; 258 AA.  
 AC Q917W7;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE Minor structural protein.  
 GN ORF3.  
 OS Norwalk-like virus NLV/White River/290/1994/US.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Caliciviridae;  
 OC Norovirus.  
 OX NCBI\_TaxID=171855;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Hu/NLV/White River/290/1994/US;  
 RX MEDLINE=97193806; PubMed=9041391;  
 RA Ando T., Monroe S.S., Noel J.S., Glass R.L.;  
 RT "A one-tube method of reverse transcription-PCR to efficiently amplify  
 RT a 3-kilobase region from the RNA polymerase gene to the poly(A) tail  
 RT of small round-structured viruses (Norwalk-like viruses).";  
 RL J. Clin. Microbiol. 35:570-577(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Hu/NLV/White River/290/1994/US;  
 RX MEDLINE=2026071; PubMed=10804147;  
 RA Ando T., Noel J.S., Fankhauser R.L.;  
 RT "Genetic classification of 'Norwalk-like viruses.'";  
 RL J. Infect. Dis. 181:S336-S348(2000).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Hu/NLV/White River/290/1994/US;  
 RA Ando T., Seto Y., Noel J.S., Monroe S.S., Glass R.L., Fankhauser R.L.;  
 RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF414423; AAL13023.1; --  
 DR InterPro; IPR004278; RNA\_capsid.  
 DR Pfam; PF03035; RNA\_capsid; 1.  
 SQ SEQUENCE 258 AA; 27501 MW; 8D7B3CF2F64E6EAB CRC64;



RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Ananides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Balow R.M., Basu A.B., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Bereman B.P., Bhandari D., Bolabakov S.,  
RA Borkova D., Borchan M.R., Bouck J., Brockstein P., Brothier P.,  
RA Burtis K.C., Busam D.A., Butler K., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablo S., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heian T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei B., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinart K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasarman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of *Drosophila melanogaster*.";  
RL Science 287:2185-2195(2000).  
RN [2]  
RN SEQUENCE FROM N.A.  
RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,  
RA Evans C.A., Gocayne J.D., Ananides P.G., Brandon R.C., Rogers Y.,  
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,  
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,  
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Drenek D., Farfan D.,  
RA Ferreira S., Frise E., Galie R.F., Garg N.S., George R.A.,  
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,  
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,  
RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,  
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
RA Stapleton M., Strong R., Svirkas R., Tector C., Tyler D.,  
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
RT "Sequencing of *Drosophila melanogaster* genome.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RN SEQUENCE FROM N.A.  
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,  
RA Tupy J.L., Bergman C., Bertram B., Carlson J.W., Celniker S.E.,  
RA Clamp M., Drysdaile R., Emmert D., Frise E., de Grey A., Harris N.,  
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,  
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,  
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;  
RT "Annotation of *Drosophila melanogaster* genome.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RN SEQUENCE FROM N.A.  
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RN SEQUENCE FROM N.A.  
RA FlyBase;

Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
-1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
EMBL; AE003632; AAF53097.2; -.  
HSP; P15822; 1BBO.  
DR FlyBase; FBgn0004579; salm.  
DR GO; GO:0016481; P:negative regulation of transcription; NAS.  
DR GO; GO:0007438; P:oocyte development; IMP.  
DR GO; GO:0007467; P:photoreceptor differentiation (sensu Drosophila); IMP.  
DR GO; GO:0045485; P:P8 differentiation; NAS.  
DR GO; GO:0007423; P:sensory organ development; IMP.  
DR InterPro; IPR007087; Znf\_C2H2.  
DR Pfam; PF00096; Znf\_C2H2; 7.  
DR SMART; SM00355; Znf\_C2H2; 7.  
DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 7.  
DR PROSITE; PS0157; ZINC\_FINGER\_C2H2\_2; 7.  
KW Metal-binding; Nuclear protein; Zinc; Zinc-finger.  
SQ SEQUENCE 1355 AA; 148979 MW; D7B05A058A1E51 CRC64;

Alignment Scores:  
Pred. No.: 0.802 Length: 1355  
Score: 91.00 Matches: 46  
Percent Similarity: 41.40% Conservative: 19  
Best Local Similarity: 29.30% Mismatches: 63  
Query Match: 10.83% Indels: 29  
DB: 5 Gaps: 7

US-10-057-510-1 (1-486) x Q9VKG2 (1-1355)  
QY 441 GGCTGANGCAGCTTTTCACAGACAGAGCCAGGACCAACAGGAGGGGCTTCT 382  
DB 1029 GlyValValAsnThrAsnProValArgProArgSerSerHisGlyHisSer 1048  
QY 381 TATCAAGT-----TCCCTTCATGCCATATGACCAAGNCTNGTGCT 337  
DB 1049 ValGlySerThrSerAlaProThrSerProSerValHisAlaSerSerGlnValIleLys 1068  
QY 336 CGAACCTTGACTGGGCAATGCTCGAGAGTCTGGGAC-----NATCCAGGG 289  
DB 1069 ArgSerSerProAlaArgSerGluAlaSerGlnGlyAlaLeuAspLeuThrProArg 1088  
QY 288 GCTGCA---TCTTCATCTCAGCCCTCGTAGCCTTCTGCCCTCTCCAGGGCTCGTCC 232  
DB 1089 AlaAlaProThrSerSerSerSerSerSerProLeuProLysGluLysProValSer 1108  
QY 231 ATTCTTCAGCTCCAGGC-----CTCTCTCCA 205  
DB 1109 ProProSerLeuPro-ArgSerProSerGlySerSerHisAlaSerAlaAsnIleLeuTh 1128  
QY 204 CACGTCCATGCTTCCACC-----TTGCCACATGGANACAGCCGTGCCCC 160  
DB 1128 rSerProLeuProProThrValGlyIleAspCysLeuProProGlyLeuGlnHis---Hi 1147  
QY 159 CACAAGCCAGGAGGACAGAGCTGGCAGAGAAAGAACAGACGATTAGACAGGGGGCAAC 100  
DB 1147 sLeuGlnGlnGlnHisGlnHisLeuMetGlnGlnGlnAlaValAlaAlaAlaAla 1167  
QY 99 AATGGAGAGGAGGAGTATCTATTAGCTTGGGCAATTTTACACCAGCAC 51  
DB 1167 aAlaGlnHisHis---HisHisGlnGlnMetAlaAlaLeuHisGlnHis 1182

RESULT 8  
Q9U412  
ID Q9U412 PRELIMINARY; PRT; 3469 AA.  
AC Q9U412  
DC 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE SANT domain protein SMRTER.  
GN SMR OR SMRTER OR CG4013.  
OS Drosophila melanogaster  
OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.

```
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9417957; PubMed=10488333;
RA Tsai C.-C., Kao H.-Y., Yao T.-P., McKeown M., Evans R.M.;
RT "SMRTER, a Drosophila nuclear receptor coregulator, reveals that Ecr-
RL mediated repression is critical for development.";
RL Mol. Cell 4:175-186(1999).
CC -|- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -|- SIMILARITY: CONTAINS 1 MYB-LIKE DOMAIN.
DR EMBL; AF175223; AAD5264.1; -.
DR Flybase; FBgn004309; Smr.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR002086; Aldehyde dehydr.
DR InterPro; IPR001005; Myb DNA binding.
DR Pfam; PF00249; myb_DNA-binding; 1.
DR SMART; SM00717; SANT; 1.
DR PROSITE; PS00687; ALDEHYDE DEHYDR_GLU; 1.
DR DNA-binding; Nuclear protein.
SQ SEQUENCE 3469 AA; 364115 MW; 6284E14C5C247CD9 CRC64;
Alignment Scores:
Pred. No.: 0.918 Length: 3469
Score: 91.00 Matches: 48
Percent Similarity: 37.04% Conservative: 22
Best Local Similarity: 25.40% Mismatches: 60
Query Match: 10.83% Indels: 60
DB: 5 Gaps: 8
US-10-057-510-1 (1-486) x Q9U412 (1-3469)
QY 479 GGANTTCCCTTTCTTCAGGCTTTTCNAAAACAGGGGNGTANGANCAGTTTCCAG 420
Db 2128 GlyGlySerSerSerGly- SerGlySerGlyProSerSer 2143
QY 419 AACAGAGGGCCAGACACACACAGGG- 393
Db 2144 AsparGHisHisGlyProProProProThrMetSerMetHisHisIleValArgSerGly 2163
QY 392 -----AGGGGGCTTTATGCAAGTTCCTTCATCCATATGATGCAAGGCTG 342
Db 2164 GlyMetTyrArgGlyAspThrValThrValPro-----SerLeuAlaAlaProSerSer 2181
QY 341 TGGCTCGAACCTTGACT-----GGCAATGCTCTGAGAGT 306
Db 2182 TyrLeuTyrProThrArgSerValIleSerIleGlyGlyValValProGlyVal 2201
QY 305 CTGGGACNATCCAGGGCTGCATCTTCATCTCAGCCCTGCGT-----AGCCTT 256
Db 2202 LeuProGlyValProGlyIleAlaLeuTyrLeuGlnProValProValProIle 2221
QY 255 CTGCCCTCTCCAGGGCTGCTTCATCTTCACCTCAGGCTCTCTCA----- 205
Db 2222 SerIleSerGlyGlnGlnLeuProProIleValGlnProProAlaGlnPro 2241
QY 204 -----CAC 202
Db 2242 ProSerGlyArgGlyValAlaValProProIleValProGlnGlnAlaHisHis 2261
QY 201 GTCCATGCTTCCACCTTGGCCATGAGNACAGCGCTGCCCAAGCCAGAGGGGACA 142
Db 2262 LeuHis-ProSer-----HisGlyHisSer-----ProSerGlnGlnGln 2276
QY 141 ACGAGCTGGCAAGAAACAGCAGTGTAGACGAGGGGCAACATGAGAGAAAGAGATC 82
Db 2276 NGInGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2296
QY 81 TATTATGAGCTGGCAATTTACAC 57
Db 2296 uValIleSerGlySerIleIleHis 2304
```

## RESULT 9

```
Q69115 PRELIMINARY; PRT; 209 AA.
AC Q69115;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE BZLF1.
OS Human herpesvirus 4 (Epstein-Barr virus).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
OX NCBI_TaxID=10376;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88306247; PubMed=2841800;
RA Jensen H.B., Miller G.;
RT "Polymorphisms of the region of the Epstein-Barr virus genome which
RT disrupts latency.";
RL Virology 185:549-564(1988).
DR EMBL; M20821; AAA65492.1; -.
SQ SEQUENCE 209 AA; 22234 MW; EDE3CE3C8E27CFE0 CRC64;
```

## Alignment Scores:

Pred. No.:	0.694	Length:	209
Score:	90.50	Matches:	33
Percent Similarity:	35.90%	Conservative:	9
Best Local Similarity:	28.21%	Mismatches:	48
Query Match:	10.77%	Indels:	27
DB:	12	Gaps:	5

US-10-057-510-1 (1-486) x Q69115 (1-209)

```
QY 422 CAGACAGAGGCCAGGACACACAGGGGGGCTTCTATGCAAGTCCCCTTCAT 363
Db 108 GlnAlaGlyGlyLualProGlnProGlyAspAsnSerThrValGln---ProAlaAla 126
QY 362 GCATATGAGCCAGCAAGNCTGTGGCTCGAACCCCTTGACTGGGCAATGCTCTGAGAGTCTG 303
Db 127 AlaVal-----ValPheAlaCysPro----- 133
QY 302 GGGACNATCCAGGGCTGCATCTT-----CATCTCAGCCCTGCGTAGCCCTTC 255
Db 134 GlyAlaAsnGlnGlyGlnGlnLeuAlaAspIleGlyAlaProGlnProAlaProAlaAla 153
QY 254 TGCCCTCTCCAGGGCTGCTTCATCTTCAGCTCCAGCGCTCCTCCACACGTCCTCATG 195
Db 154 AlaProAlaArgArgThrArgLysProLeuGln----- 164
QY 194 CTTTCCACCTTCCACATGGANACAGCCGTGCCCCCAAGCCAGAGGGAACAGAAAGC 135
Db 165 ProGluSerValArgArgGlySerLysArgValAlaAlaAsnGlyGlyAlaGlnArgGly 184
QY 134 TGGCAAGAAAGAACAGCAGTGTAGACGAGGGCAACAATGGAGAAAGAGTA 84
Db 185 TrpGlnArgThrValGlyLeuLysGluGlySerGlyArgTrpGlyLeu 201
```

## RESULT 10

```
Q69123 PRELIMINARY; PRT; 209 AA.
AC Q69123;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE BZLF1.
OS Human herpesvirus 4 (Epstein-Barr virus).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
OX NCBI_TaxID=10376;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88306247; PubMed=2841800;
RA Jensen H.B., Miller G.;
```

RT "Polymorphisms of the region of the Epstein-Barr virus genome which  
 RT disrupts latency";  
 RL Virology 185:549-564 (1988).  
 DR EMBL: M20830; AAA65495.1; --  
 SQ SEQUENCE 209 AA; 22278 MW; 876D395F478CF59 CRC64;

Alignment Scores:  
 Pred. No.: 0.694 Length: 209  
 Score: 90.50 Matches: 32  
 Percent Similarity: 35.04% Conservative: 9  
 Best Local Similarity: 27.35% Mismatches: 49  
 Query Match: 10.77% Indels: 27  
 DB: 12 Gaps: 4

US-10-057-510-1 (1-486) x Q69123 (1-209)

QY 422 CAGAACAGAGCCAGGACACAGCAGGAGGGGCTTCATTATGCAAGTCCCTTCAT 363  
 DB 108 GlnAlaGlyGlyGluAlaProGlnProGlyAspAsnSerThrValGlnProGluAla 127  
 QY 362 GCCATATGAGCCCAAGNCCTNGTGGCTCGAACCCCTTGACTGGGCAATGCTCGAGAGTCTG 303  
 DB 128 ValValPheAla-----CysPro----- 133  
 QY 302 GGGACNATCCAGGGCTGCATCTT-----CATCTCAGCCCTGCGTAGCCCTTC 255  
 DB 134 GlyAlaAsnGlnGlyGlnGlnLeuAlaAspIleGlyAlaProGlnProAlaProAla 153  
 QY 254 TGCCTCTCTCCAGGGCTCGTCATCTTCCTCAGCTCCAGCGCTCTCCACAGCTCCATG 195  
 DB 154 AlaProAlaArgArgThrArgLysProLeuGln----- 164  
 QY 194 CTTCCACCTTCCACATGGANACAGCGTGGTCCGCCACAGCAGGAGGACAAACGAGC 135  
 DB 165 ProGluSerValArgArgLysArgValAlaAlaAsnGlyGlyAlaGlnArgGly 184  
 QY 134 TGCAGAGAAAGAACAGCAGCTTAGAGCAGGGGCAACAATGGAGAAAGGAGTA 84  
 DB 185 TrpGlnArgThrValGlyLeuLysGluGlySerGlyArgTrpGlyLeu 201

RESULT 11

QXZN8 PRELIMINARY; PRT; 647 AA.  
 AC Q8XZN8;  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 23, Last annotation update)  
 DE GALA protein 5.  
 GN RSC1357 OR RS04639.  
 OS Ralstonia solanacearum (Pseudomonas solanacearum).  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 OC Burkholderiaceae; Ralstonia.  
 OC NCBI\_TaxID=305;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GM11000;  
 RX MEDLINE=21681879; PubMed=11823852;  
 RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,  
 RA Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,  
 RA Chandler C., Choise N., Claudei-Renard C., Cunnac S., Demange N.,  
 RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,  
 RA Sigulier P., Thebaud P., Whalen M., Wincker P., Levy M.,  
 RA Weissenbach J., Boucher C.A.;  
 RT "Genome sequence of the plant pathogen Ralstonia solanacearum";  
 RL Nature 415:497-502 (2002).  
 DR EMBL; AL646064; CND15059.1; --  
 DR InterPro; IPR001810; F-box.  
 DR InterPro; IPR001611; LRR.  
 DR Pfam; PF00646; F-box; 1.  
 DR Pfam; PF00560; LRR; 2.  
 DR SMART; SM00256; FBOX; 1.  
 DR PROSITE; PS0181; FBOX; 1.  
 KW Complete proteome.

SQ SEQUENCE 647 AA; 68312 MW; 8880B499A26F0F25 CRC64;

Alignment Scores:  
 Pred. No.: 1.05 Length: 647  
 Score: 89.50 Matches: 42  
 Percent Similarity: 38.71% Conservative: 18  
 Best Local Similarity: 27.10% Mismatches: 57  
 Query Match: 10.65% Indels: 38  
 DB: 16 Gaps: 5

US-10-057-510-1 (1-486) x Q8XZN8 (1-647)

QY 486 CCCATGGGANTTCCCTTCTTTCAGGCTTTCNAAAACAGGGGNCCTGANGCAGT 427  
 DB 53 ProGluGlyGlyPheProHis-----ValProAspGln 63  
 QY 426 TTTCAGAACAGAGGGCCAGGACACAAACAGGAGGGGGCTTCTTATGCAAGTCCCT 367  
 DB 64 AlaGlyGluGlnArgAlaTySerProSerArgSerAlaGlyValSerSerPro 83  
 QY 366 TCATGCCATATGAGCCCAAGNCCTNGTGGCTCGAACCCCTTGACTGGGCAATGCTCGAGAG 307  
 DB 84 -----LeuGlyGlyLeuAlaAlaMetArgLeuAsp----- 93  
 QY 306 TCTGGGACNATCCAGGGCTGCATCTTCTCAGCCCTGCGTAGCTTCTGCGCTCC 247  
 DB 94 -----SerGlySerAlaGlyThrAlaProMetLeuArgSerArgLeuPro 109  
 QY 246 TCCAGGGCTCGTCCATCTCAGCTCCAGCGCTCTCCACAGTCCATGCTTC-- 190  
 DB 110 Ala-----ProSerAlaProSerGlyValProGlnValArgAsnIleGlu 124  
 QY 189 -----CAGCTTGCACATGGANACAGCGCTGCGCCCAACAG 154  
 DB 125 GluMetProAlaGlyValLeuGlnHisValAlaSerPheLeuAspProArgSerArg 144  
 QY 153 CCAGAGGGAACAGCAGAGCTGGCAGAAAGAACAGCAGATTAGACGAGGGGCAACAATGGA 94  
 DB 145 -AlaLeuSerGlnValSerThrThrMetAsnAspAlaAlaArgSerGlnThrHisMe 164

RESULT 12

Q9HCL7 PRELIMINARY; PRT; 2414 AA.  
 AC Q9HCL7;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein KIAA1555 (Fragment).  
 GN KIAA1555.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=20450693; PubMed=10997877;  
 RA Nagase T., Kikuno R., Nakayama M., Hirose M., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes.  
 RT XVIII. The complete sequences of 100 new cDNA clones from brain which  
 RT code for large proteins in vitro";  
 RL DNA Res. 7:273-281 (2000).  
 DR EMBL; AB046775; BAB13381.2; --  
 DR GO; GO:0016829; F-lyase activity; IEA.  
 DR GO; GO:0008152; P-metabolism; IEA.  
 DR InterPro; IPR002034; AIPM/HcIt\_synth.  
 DR InterPro; IPR007087; Znf.C2H2.  
 DR Pfam; PF00096; zf-C2H2; 2.  
 DR SMART; SM00355; Znf\_C2H2; 5.







RT avermitilis: deducing the ability of producing secondary  
 RT metabolites";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).  
 RN [2]

RP SEQUENCE FROM N.A.  
 RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;  
 RX MEDLINE=22608306; PubMed=12692562;  
 RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,  
 RT Sakaki Y., Hattori M., Omura S.;  
 RA "Complete genome sequence and comparative analysis of the industrial  
 RT microorganism Streptomyces avermitilis";  
 RL Nat. Biotechnol. 21:526-531(2003).  
 DR EMBL: AP005029; BAC69919.1; -;  
 DR InterPro: IPR005654; AFG1\_ATPase.  
 DR Pfam: PF03969; AFG1\_ATPase; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 378 AA; 40246 MW; BCD53E5ADAFE2356 CRC64;

## Alignment Scores:

Pred. No.:	2.3	Length:	378
Score:	86.00	Matches:	37
Percent Similarity:	35.33%	Conservative:	16
Best Local Similarity:	24.67%	Mismatches:	37
Query Match:	10.24%	Indels:	60
DB:	16	Gaps:	7

US-10-057-510-1 (1-486) x Q82L04 (1-378)

QY	342	GTGGCTCGAACCTTGACTGGCAATGCTCTGAGAGTCITGGGACNATCCAGGGGCTGCA	283
Db	5	VaAlaAlaArgInLeuGluTrp-----GlyThrVal	14
QY	282	TCITCATCTCCAGCCGTGCGTAGCCTTCCTCCCTCCAGGGCTCGTCCATT-----	229
Db	15	SerSerSerThrAlaAlaSerGlyIleAspProIleAlaGluAlaProLeuSerLeu	34
QY	228	-----CCTTCAGCTCCAGCG-----CCTCTCCA	205
Db	35	CysAlaAlaArgAlaProHisValProAlaAspArgLeuValAlaGluMetValProPro	54
QY	204	CACGTCCATGCTCCACCTTGCCACATGGANACAGCGCGTCCACACAGCCAGCCAGGGA	145
Db	55	ArgPheAspSerValArgPheAlaThrTyIle---ProAspProAsnGlnProSerGln	73
QY	144	ACA-----ACGAAGCTGGCAAGAAAGACAGCAGT	115
Db	74	ThrGluAlaValArgValLeuGluAspPheAlaThrGlyLeuGlyGlyAlaHisAlaSer	93
QY	114	TAGAGC-----AGGGCAACAAATGGAGAA	91
Db	94	GlySerGlyArgArgArgLeuPheGlyPheGlyLysAlaProLysThrProAlaGlyPro	113
QY	90	AGGAGTATCTATTATGAG-----CTT	70
Db	114	ArgGlyValTyLeuAspGlyGlyTyGlyValGlyLysThrHisLeuLeuAlaSerLeu	133
QY	69	GGGCAATTTACCCAGCACTTCCAGCCTT	40
Db	134	TrpHisAlaThrProAlaGluProGluLeu	143

Search completed: April 6, 2004, 17:57:46  
 Job time : 51.5 secs



```
FT TRANSMEM 95 115 2 (POTENTIAL).
FT DOMAIN 116 125 3 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 126 146 3 (POTENTIAL).
FT DOMAIN 147 165 4 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 166 186 4 (POTENTIAL).
FT DOMAIN 187 206 5 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 207 227 5 (POTENTIAL).
FT DOMAIN 228 243 6 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 244 264 6 (POTENTIAL).
FT DOMAIN 265 286 7 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 287 307 7 (POTENTIAL).
FT DOMAIN 308 335 7 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 16 16 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 26 26 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 32 32 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 335 AA; 35377 MW; D0FFB02DBB9A17D7 CRC64;

Alignment Scores:
Pred. No.: 2.07 Length: 335
Score: 88.00 Matches: 56
Percent Similarity: 38.50% Conservatve: 16
Best Local Similarity: 29.95% Mismatches: 55
Query Match: 10.05% Indels: 60
DB: 1 Gaps: 12

US-10-057-510-1 (1-486) x DUFF_MACMU (1-335)

Qy 42 GGCTGGGAAGTGGTGTAAATGCCCAAGCTCATATAGATACCTCT--TTCTCCAT 98
Db 94 GlyTrpProValuu-----AlaGln-LeuAlaValGlySerAlaLeuPheSerI 110
Qy 99 TGTTCGCCCTGCTTAACCTGCTGTTCTTCT-----TGCCAGCT 137
Db 110 eValValProIleLeuAlaProGlyLeuAlaThrArgSerSerAlaLeuCysSerLe 130
Qy 138 TCGTGTGCTCCTCGG-----CTGTGGGG-----GC 164
Db 130 uGlyTyrcysValTrpTrpGlySerAlaPheAlaGlnAlaLeuLeuLeuGlyCysHisal 150
Qy 165 ACGGCTGTNTCCCA-----TGTGGCAAGGTGGAAGGATGAGCTGTGGAGGAGGG 215
Db 150 aSerLeuGlyProIleLeuGlyAlaGlnValProGlyLeu-ThrLeuGlyLeuSerV 170
Qy 216 CTGAGCTGAAGG-----ATGACAGAGCCTGGGAGGAGGAGGAGGAGGAGCTAC 263
Db 170 alGlyLeuTrpGlyValAlaAlaLeuLeu-ThrLeuProIleThrLeuAlaSerGlyAla 189
Qy 264 GCAGGG-----CTGAGGATGAAGATGAGCCCTCGATNGTCCCCAG 305
Db 190 SerGlyLeuCysThrProAlaTyrcysMetGluLeuLeuAlaGlnAlaThrHis 209
Qy 306 ACTCTCAGGACATGTCACATCAAGGTTTCGAGCCACNAGNCTT---GGCTCATATGGC 362
Db 210 AlaValAlaCysLeuAlaValPheValLeuLeuProLeuGlyLeuPheGlyAlaLysGly 229
Qy 363 ATGAGGGGAACCTGTCATAGAGCCCTCCCTCGG-----TTGTGGTCTCGGCC 413
Db 230 LeuLysLysAlaLeuGlyMetGlyProGlyProIleMetAsnIleLeuIlePheAlaTrp-Phe 249
Qy 414 TCTGTTCTGGAACACTGGTTCNTAGNCCCTCGTTTTCGCAAGACCTGGAAGAAAGGA 473
Db 249 eIlePheTrp-----Trp-----253

Qy 474 AANTCCCATGGG 486
Db 254 ----ProHisGly 256

RESULT 2
SALM_DROME STANDARD; PRT; 1355 AA.
AC P39770;
DT 01-FEB-1995 (Rel. 31, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
```

```
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Homeotic protein spalt-major.
GN SALM.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94139659; PubMed=7905822;
RA Kuehnlein R.P., Frommer G., Friedrich M., Gonzalez-Gaitan M.,
RA Weber A., Wagner-Bernholz J.F., Gehring W.J., Jaeckle H., Schuh R.;
RT "Spalt encodes an evolutionarily conserved zinc finger protein of
RT novel structure which provides homeotic gene function in the head and
RT tail region of the Drosophila embryo.";
RL EMBO J. 13:168-179(1994).
CC -!- FUNCTION: REQUIRED FOR THE ESTABLISHMENT OF THE POSTERIOR-MOST
CC HEAD AND THE ANTERIOR-MOST TAIL SEGMENTS OF THE EMBRYO. PROBABLY
CC FUNCTION AS A TRANSCRIPTIONAL REGULATOR. COULD REPRESS THE
CC TRANSCRIPTION OF THE TSH GENE.
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- DEVELOPMENTAL STAGE: FIRST EXPRESSED AT BLASTODERM STAGE AND
CC LATER IN RESTRICTED AREAS OF THE EMBRYONIC NERVOUS SYSTEM AS
CC WELL AS IN THE DEVELOPING TRACHEA.
CC -!- SIMILARITY: BELONGS TO THE SAL FAMILY OF C2H2-TYPE ZINC-FINGER
CC PROTEINS.
CC -!- SIMILARITY: Contains 7 C2H2-type zinc fingers.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X75541; CAA53229.1; -.
CC PIR; S40022; S40022.
CC DR HSP; P15822; 1BEO.
CC DR FlyBase; FBgn0004579; salm.
CC DR GO; GO:0016481; P:negative regulation of transcription; NAS.
CC DR GO; GO:0007438; P:cytocyte development; IMP.
CC DR GO; GO:0007487; P:photoreceptor differentiation (sensu Drosop. . .); IMP.
CC DR GO; GO:0045485; P:R8 differentiation; NAS.
CC DR GO; GO:0007423; P:sensory organ development; IMP.
CC DR InterPro; IPR007087; Znf_C2H2.
CC DR Pfam; PF00096; zf-C2H2; 7.
CC DR SMART; SM00355; Znf_C2H2; 7.
CC DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 7.
CC DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 7.
CC DR Developmental protein; Zinc-finger; Metal-binding; DNA-binding;
CC Nuclear protein; Transcription regulation; Repeat.
CC ZN_FING 451 473 C2H2-TYPE 1.
FT ZN_FING 479 501 C2H2-TYPE 2.
FT ZN_FING 824 846 C2H2-TYPE 3.
FT ZN_FING 852 874 C2H2-TYPE 4.
FT ZN_FING 884 906 C2H2-TYPE 5.
FT ZN_FING 1289 1311 C2H2-TYPE 6.
FT ZN_FING 1317 1339 C2H2-TYPE 7.
SQ SEQUENCE 1355 AA; 148995 MW; 6FC4EBDC0BC6355E CRC64;

Alignment Scores:
Pred. No.: 4.82 Length: 1355
Score: 84.50 Matches: 46
Percent Similarity: 39.29% Conservatve: 20
Best Local Similarity: 27.38% Mismatches: 56
Query Match: 10.06% Indels: 36
DB: 1 Gaps: 7

US-10-057-510-1 (1-486) x SALM_DROME (1-1355)
Qy 441 GGNCTGANGNCAGTTTTCAGACAGAGGCCAGGCCACACACAGGAGGGGCTTCT 382
```

```

Db      1029 GlyValValAsnThrAsnProValArgSerSerAlaSerSerHisGlyHisSer 1048
QY      381  TATGCAAGT-----TCCCTTCATCCATATGACCAAGNCTGTGCT 337
Db      1049 ValGlySerThrSerAlaProThrSerProSerValHisAlaSerSerGlnValIleLys 1068
QY      336  CCAACCCCTGACTGGGCATGCTCAGAGTCTGGGAC-----NATCCAGGG 289
Db      1069 ArgSerSerSerProAlaArgSerGlnAlaSerGlnGlyAlaLeuAspLeuThrProArg 1088
QY      288  GTGCA---TCTTCATCTCCAGCCCTGCGTAGCCCTTCTGCCCTCCACAGGCTCGTCC 232
Db      1089 AlaAlaProThrSerSerSerSerArgSerProLeuProLysGluLysProValSer 1108
QY      231  ATTCCTTCAGTCCAGCCG-----TTGCCACATGGANACAGCGGTGCC 205
Db      1109 ProProSerLeuProArgSerProSerGlySerHisAlaSerAlaAsnIleLeuTh 1128
QY      204  CACGTCCATGCTTCCACC-----TTGCCACATGGANACAGCGGTGCC 160
Db      1128 rSerProLeuProProThrValGlyLeaspCysLeuProLysGlyLeuGlnHis---Hi 1147
QY      159  CACAGCCAGAGGGAACAGAGCTGGCAAGAAAGACAGCAGTAGTAGAGGCGGCAAC 100
Db      1147 sleuGlnGlnGlnHisGlnHisLeuMetGlnGlnGlnAlaValAlaAlaAlaAla 1167
QY      99  AATGGAGAAGAGTATCTATTATGAGCTGGGCATTTTACACAGCAGCAGCTCCAGCCTT 40
Db      1167 alaGlnHisHis-----HisHisGlnGlnMetAlaAlaLe 1179
QY      39  TTCCATAACATGGAAACATT 18
Db      1179 uAspGlnHisGlnGlnGlnLeu 1186

RESULT 3
DUFF_HUMAN
ID  DUFF_HUMAN          STANDARD;          PRT;          336 AA.
AC  Q16570; O75898; Q16300; Q8WME3; Q9UUF0; Q9UKZ5; Q9UKZ6; Q9UOE1;
DT  28-FEB-2003 (Rel. 41, Last sequence update)
DT  10-OCT-2003 (Rel. 42, Last annotation update)
DE  Duffy antigen/chemokine receptor (Fy glycoprotein) (GpFy)
GN  Fy OR DARC OR GDP.
OS  Homo sapiens (Human).
OC  Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX  NCBI_TaxID=9606;
RN  [1]_
SEQUENCE FROM N.A. (ISOFORM 1), AND PARTIAL SEQUENCE.
RP  TISSUE=Blood;
RX  MEDLINE=94068488; PubMed=8248172;
RA  Chaudhuri A., Polyakova J., Zbrzezna V., Williams K., Gulati S.,
RA  Pogo A.;
RT  "Cloning of glycoprotein D cDNA, which encodes the major subunit of
RT  the Duffy blood group system and the receptor for the Plasmodium
RT  vivax malaria parasite.";
RL  Proc. Natl. Acad. Sci. U.S.A. 90:10793-10797(1993).
RN  [2]
SEQUENCE FROM N.A. (ISOFORM 1).
RP  TISSUE=Blood;
RX  MEDLINE=95392577; PubMed=7663520;
RA  Tournamille C., Colin Y., Carttron J.-P., Le van Kim C.;
RT  "Disruption of a GATA motif in the Duffy gene promoter abolishes
RT  erythroid gene expression in Duffy-negative individuals.";
RL  Nat. Genet. 10:224-228(1995).
RN  [3]
SEQUENCE FROM N.A. (ISOFORM 1), AND VARIANT GLX-42.
RP  TISSUE=Blood;
RX  MEDLINE=95134891; PubMed=7833467;
RA  Iwamoto S., Omi T., Kajii E., Ikemoto S.;
RT  "Genomic organization of the glycoprotein D gene: Duffy blood group

```

```

RT  Fya/Fyb alloantigen system is associated with a polymorphism at the
RT  44-amino acid residue.";
RL  Blood 85:622-626(1995).
RN  [4]
SEQUENCE FROM N.A. (ISOFORM 2), AND VARIANTS CYS-89 AND THR-100.
RP  TISSUE=Blood;
RX  MEDLINE=98402363; PubMed=9731074;
RA  Tournamille C., Le Van Kim C., Gane P., Le Pennec P.Y., Roubinet F.,
RA  Babinet J., Carttron J.-P., Colin Y.;
RT  "Arg89Cys substitution results in very low membrane expression of the
RT  Duffy antigen/receptor for chemokines in Fy(x) individuals.";
RL  Blood 92:2147-2156(1998).
RN  [5]
SEQUENCE FROM N.A. (ISOFORM 1), AND VARIANTS CYS-89 AND THR-100.
RP  MEDLINE=99101431; PubMed=9886340;
RA  Olsson M.L., Smythe J.S., Hansson C., Poole J., Mallinson G.,
RA  Jones J., Arent N.D., Daniels G.;
RT  "The Fy(x) phenotype is associated with a missense mutation in the
RT  Fy(b) allele predicting Arg89Cys in the Duffy glycoprotein.";
RL  Br. J. Haematol. 103:1184-1191(1998).
RN  [6]
SEQUENCE FROM N.A. (ISOFORM 2), AND VARIANT GLY-42.
RP  Begguley C.;
RL  Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
RN  [7]
SEQUENCE FROM N.A. (ISOFORM 2), AND VARIANT GLY-42.
RP  TISSUE=Lung;
RX  MEDLINE=22388257; PubMed=12477932;
RA  Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA  Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA  Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA  Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA  Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA  Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA  Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA  Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA  Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA  Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA  Vallalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA  Fahey J., Helton E., Kettaman M., Madan A., Rodriguez S., Sanchez A.,
RA  Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA  Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA  Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA  Butterfield J.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,
RA  Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT  "Generation and initial analysis of more than 15,000 full-length
RT  human and mouse cDNA sequences.";
RN  Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN  [8]
SEQUENCE OF 1-117 FROM N.A. (ISOFORMS 1 AND 2), AND VARIANT GLY-42.
RX  MEDLINE=20040661; PubMed=10570183;
RA  Zimmerman P.A., Woolley I., Masinde G.L., Miller S.M., McNamara D.T.,
RA  Hazlett F., Mgone C.S., Alpers M.P., Genton B., Boatín B.A.,
RA  Kazura J.W.;
RT  "Emergence of FyA*(null) in a Plasmodium vivax-endemic region of
RT  Papua New Guinea.";
RL  Proc. Natl. Acad. Sci. U.S.A. 96:13973-13977(1999).
RN  [9]
VARIANT GLY-42.
RP  TISSUE=Peripherel blood;
RX  MEDLINE=95220914; PubMed=7705836;
RA  Tournamille C., Le van Kim C., Gane P., Carttron J.-P., Colin Y.;
RT  "Molecular basis and PCR-DNA typing of the Fya/fyb blood group
RT  polymorphism.";
RL  Hum. Genet. 95:407-410(1995).
RN  [10]
VARIANT CYS-89.
RX  MEDLINE=98421365; PubMed=9746760;
RA  Parasol N., Reid M., Rios M., Castilho L., Harari I., Kosower N.S.;
RT  "A novel mutation in the coding sequence of the Fy* allele of the
RT  Duffy chemokine receptor gene is associated with an altered
RT  erythrocyte phenotype.";
RL  Blood 92:2237-2243(1998).

```

CC -!- FUNCTION: Non-specific receptor for many chemokines such as IL-8,  
 CC GRO, RANTES, MCP-1 and TARC. It is also the receptor for the human  
 CC malaria parasites Plasmodium vivax and Plasmodium knowlesi.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=2;  
 CC IsoId=Q16570-1; Sequence=Displayed;  
 CC Name=1;  
 CC IsoId=Q16570-2; Sequence=VSP\_001323;  
 CC -!- TISSUE SPECIFICITY: Found in adult kidney, adult spleen, bone  
 CC marrow and fetal liver. In particular, it is expressed along  
 CC postcapillary venules throughout the body, except in the adult  
 CC liver. Erythroid cells and postcapillary venule endothelium are  
 CC the principle tissues expressing duffy. Fy(-A-B) individuals do  
 CC not express duffy in the bone marrow, however they do, in  
 CC postcapillary venule endothelium.  
 CC -!- POLYMORPHISM: DY is responsible for the Duffy blood group system.  
 CC The molecular basis of the Fy(A)=Fy1/Fy(B)=Fy2 blood group  
 CC antigens is a single variation in position 42; Gly-42 corresponds  
 CC to Fy(A) and Asp-42 to Fy(B).  
 CC -!- DISEASE: Individuals that do not produce the Duffy antigen  
 CC (Fy(A-B-)) are more resistant to vivax malaria. This allele is  
 CC found predominantly in population of african origin.  
 CC -!- SIMILARITY: Belongs to family Duffy of G-protein coupled  
 CC receptors.  
 CC -!- DATABASE: NAME=PROW; NOTE=PROW 2:90-94(2001);  
 CC WWW='http://www.ncbi.nlm.nih.gov/prow/guide/303294872.g.htm'.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL; U01839; AAC50055.1; -  
 CC DR EMBL; X85785; CAAS9770.1; -  
 CC DR EMBL; S76830; AAB33239.1; -  
 CC DR EMBL; AF055992; AAC72301.1; -  
 CC DR EMBL; AF030521; AAD20435.1; -  
 CC DR EMBL; AL035403; CAB56228.1; -  
 CC DR EMBL; BC017817; AAF17817.1; -  
 CC DR EMBL; AF100634; AAF02415.1; -  
 CC DR EMBL; AF100634; AAF02416.1; -  
 CC DR PIR; I52608; I52608.  
 CC DR Genew; HGNC:4035; FY.  
 CC DR MIM; 110700; -  
 CC DR GO; GO:0005886; C:plasma membrane; TAS.  
 CC DR GO; GO:0004888; F:transmembrane receptor activity; TAS.  
 CC DR InterPro; IPR005384; Duffyantigen.  
 CC DR PRINTS; PR01559; DUFFYANTIGEN.  
 CC KW G-protein coupled receptor; Transmembrane; Glycoprotein;  
 CC Alternative splicing; Polymorphism; Blood group antigen.  
 CC FT DOMAIN 1 63  
 CC FT TRANSMEM 64 84  
 CC FT DOMAIN 85 95  
 CC FT TRANSMEM 96 116  
 CC FT DOMAIN 117 126  
 CC FT TRANSMEM 127 147  
 CC FT DOMAIN 148 166  
 CC FT TRANSMEM 167 187  
 CC FT DOMAIN 188 207  
 CC FT TRANSMEM 208 228  
 CC FT DOMAIN 229 244  
 CC FT TRANSMEM 245 265  
 CC FT DOMAIN 266 287  
 CC FT TRANSMEM 288 308  
 CC FT DOMAIN 309 336  
 CC FT CARBOHYD 16 16  
 CC FT CARBOHYD 27 27  
 CC FT CARBOHYD 33 33

FT VARSPLIC 1 7 MGNCLHR -> MASSGVVLQ (in isoform 1).  
 FT FT FTID=VSP\_001323.  
 FT VARIANT 42 42 D -> G (in Fy(a) antigen; dbSNP:12075).  
 FT FT FTID=VAR\_003480.  
 FT VARIANT 89 89 R -> C (in Fy(x)).  
 FT FT FTID=VAR\_015068.  
 FT VARIANT 100 100 A -> T (in dbSNP:13962).  
 FT FT FTID=VAR\_015069.  
 FT CONFLICT 326 326 S -> F (in REF. 7).  
 SQ SEQUENCE 336 AA; 35610 MW; ACCEA15FBFICBE9B CRC64;  
 Alignment Scores:  
 Pred No.: 4.82 Length: 336  
 Scores: 84.00 Matches: 54  
 Percent Similarity: 37.23% Conservative: 16  
 Best Local Similarity: 28.72% Mismatches: 56  
 Query Match: 9.59% Indels: 62  
 DB: 1 Gaps: 11  
 US-10-057-510-1 (1-486) x DUFF\_HUMAN (1-336)  
 QY 42 GCGTGGGAAGTCTGCTGTAATAATGCCCAAGCTCATATAGATACCTCT---TTCTCCAT 98  
 Db 95 GlyTrpProValLeu-----AlaGln-LeuAlaValGlySerAlaLeuPheSerIl 111  
 QY 99 TGTTCGCCCTGCTCTAACTGCTGTTCTTCT-----TGCCAGCT 137  
 Db 111 eValValProValLeuAlaProGlyLeuGlySerThrArgSerSerAlaLeuCysSerLe 131  
 QY 138 TCGTTGTTCCCTCTGGCTT----- 156  
 Db 131 uGlyTyxCysValTrpTyrglySerAlaPheAlaGlnAlaLeuLeuGlyCysHisAl 151  
 QY 157 ---GTGGGSCACGGTGTNTCCATGTCGCAAGTGGAAAGCATGACGCTGGAGGAG 212  
 Db 151 aSerLeuGlyHisArgLeuGly---AlaGlyGlnValProGlyLeu-ThrLeuGlyLeu 170  
 QY 213 GCGTGGAGCTGAAGGA-----ATGGACAGACCCCTGGGAGGAGGCGAGAGCC 260  
 Db 170 hrValGlyIleTrpGlyValAlaAlaLeuLeu-ThrLeuProValThrLeuAlaSerGly 189  
 QY 261 TACGACGG-----CTGAGGATGAAGATGACGACCCCTGGATNGTCCC 302  
 Db 190 AlaSerGlyGlyLeuCysThrLeuIleTy:SerThrGluLeuGlyAlaLeuGlnAlaThr 209  
 QY 303 CAGACTCTCAGGACATTCGCCAGTCAAGGTTGCGAGCCACNAGNCCTT---GGCTCATAT 359  
 Db 210 HisThrValAlaCysLeuAlaIlePheValLeuLeuProLeuGlyLeuPheGlyAlaLys 229  
 QY 360 GGCATGAAGGGAACCTTGATCAAGAACCCCTCCCTGG-----TTGTGCTCCTGG 410  
 Db 230 GlyLeuGlyLysAlaLeuGlyMetGlyProGlyProTrpMetAlaIleLeuTrpAlaTrp 249  
 QY 411 CCCTCTGTTCGTGAAGAACTGNTCNTAGNCCCTCGTGTGTTTTTNGCAAAGCCTGAAGAAG 470  
 Db 250 -PheIlePheTrp-----Tip----- 254  
 QY 471 GGAAATCCCATCGG 486  
 Db 255 -----ProHisGly 257  
 RESULT 4  
 GDAL WHEAT  
 ID GDAL WHEAT STANDARD; PRT; 262 AA.  
 AC P04721;  
 DT 13-AUG-1987 (Rel. 05, Created)  
 DT 13-AUG-1987 (Rel. 05, last sequence update)  
 DT 15-JUL-1999 (Rel. 38, last annotation update)  
 DS Alpha/beta-gliadin A-1 precursor (Prolamin).  
 OS Triticum aestivum (Wheat).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;  
 OC Triticaceae; Triticum.

```

OX NCBI_TaxID=4565;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85234522; PubMed=2989281;
RA Okita T.W., Cheesbrough V., Reeves C.D.;
RT "Evolution and heterogeneity of the alpha-/beta-type and gamma-type
RT gliadin DNA sequences";
RL J. Biol. Chem. 260:8203-8213 (1985).
CC -!- FUNCTION: Gliadin is the major seed storage protein in wheat.
CC -!- MISCELLANEOUS: The alpha/beta-gliadins can be divided into 5
CC homology classes. Sequence divergence between the classes is due
CC to single base substitutions and to duplications or deletions
CC within or near direct repeats. There are more than a 100 copies of
CC the gene for alpha/beta-gliadin per haploid genome.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M11074; AAA34281.1; -.
DR PIR; B22364; B22364.
DR InterPro; IPR003612; AA1.
DR DRIPrints; PR00208; GLIADGLUTEN.
DR SMART; SM00499; AA1; 1.
KW Seed storage protein; Repeat; Signal; Multigene family.
FT SIGNAL 1 20
FT CHAIN 21 262 ALPHA/BETA-GLIADIN A-1.
SQ SEQUENCE 262 AA; 30403 MW; 225C8A80B6D6E3B5 CRC64;

Alignment Scores:
Pred. No.: 5.26 Length: 262
Score: 83.50 Matches: 34
Percent Similarity: 39.37% Conservative: 16
Best Local Similarity: 26.77% Mismatches: 43
Query Match: 9.94% Indels: 34
DB: 1 Gaps: 4

US-10-057-510-1 (1-486) x GDAL_WHEAT (1-262)
QY 431 CCAGTTTTCAGAACAGAGGGCCAGGACACACACAGGAGGGGGCTTCTTATGCAAGTT 372
Db 26 ProGlnLeuGlnProGlnAsnProSerGlnGlnProGln-----GlnGlnVal 42
QY 371 CCCCTTCATGCCATATAGCCAGGACGCTGTGGTGCAGACCTTGACTGGGCAATGCTCT 312
Db 43 ProLeuValGlnGlnGlnGlnPheLeuGlyGlnGlnProPheProGlnGlnPro 62
QY 311 GAGAGTCTGGGACNATCCAGGGGCTGCATCTTCATCTCCAGCCCTGCGTAGCCTTCTGC 252
Db 63 -----TyrProGlnPro-----Gln 67
QY 251 CTCTCTCCAGGGCTCGTCATCTCTAGCTCCAGCGCTCTCTCCACAGTCCTAGCTCT 192
Db 68 ProPheProSerGlnGlnProTyrLeuGlnLeuGlnProPheLeuGlnProGlnLeuPro 87
QY 191 TCACCTCTGCCA-----CAT 177
Db 88 TyrSerGlnProGlnProPheArgProGlnGlnProTyrProGlnProGlnProGlnTyr 107
QY 176 GGANACAGCGCTGCCCCCAAGCCAGAGGGAACACAGAGCTGGCAAGAAAGACAGCA 117
Db 108 SerGlnProGlnProPheLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 127
QY 116 GTTAGAGCGGGGACAA 98
Db 127 nGlnGlnGlnGlnGlnGln 133

```

```

RESULT 5
VRP1_YEAST
ID VRP1_YEAST STANDARD; PRT; 817 AA.
AC P37370; Q06133;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Verprolin.
GN VRP1 OR MDP2 OR ENDS OR YLR337W OR L8300.13.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A364;
RX MEDLINE=95058201; PubMed=7968536;
RA Donnelly S.F.H., Pocklington M.J., Pallota D., Orr E.;
RT "A proline-rich protein, verprolin, involved in cytoskeletal
RT organization and cellular growth in the yeast Saccharomyces
RT cerevisiae.";
RL Mol. Microbiol. 10:585-596 (1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=97313267; PubMed=9169871;
RA Johnston M., Hillier L., Riles L., Albermann K., Andre B., Ansoerge W.,
RA Benes V., Brueckner M., Delius H., Dubois E., Duesterhoeft A.,
RA Enrian K.-D., Floeth W., Goffeau A., Hebling U., Heumann K.,
RA Heuss-Neitzel D., Hilbert H., Hilger F., Kleine K., Koetter P.,
RA Louis E.J., Messenguy P., Mewes H.-W., Mioega T., Moestl D.,
RA Mueller-Auer S., Nentwich U., Obermaier B., Piravandi E., Pohl T.M.,
RA Portetle D., Purnelle B., Rechmann S., Rieger M., Rinke M., Rose M.,
RA Scharte M., Scherens B., Scholler P., Schwager C., Schwarz S.,
RA Underwood A.P., Urrestarazu L.A., Vandenbol M., Verhasselt P.,
RA Vierendeels F., Voet M., Volckaert G., Voss H., Wambutt R., Wedler E.,
RA Wedler H., Zimmermann F.K., Zollner A., Hani J., Honeisel J.D.; XII.";
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XII.";
RL Nature 387:87-90 (1997).
CC -!- FUNCTION: Involved in cytoskeletal organization and cellular
CC growth. May exert its effects on the cytoskeleton directly, or
CC indirectly via proline-binding proteins (e.g. profilin) or
CC proteins possessing SH3 domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z26545; CAAB1388.1; -.
DR EMBL; U19028; AAB67263.1; -.
DR PIR; S51342; S51342.
DR GeneOnline; 142401; -.
DR SGD; S0004328; VRP1.
DR GO; GO:0005857; C:actin cortical patch (sensu Saccharomycetes); IDA.
DR GO; GO:0003779; F:actin binding; IDA.
DR GO; GO:0007015; F:actin filament organization; IPI.
DR GO; GO:0006897; P:endocytosis; IMP.
DR GO; GO:0007121; P:polar budding; IMP.
DR GO; GO:0006970; P:response to osmotic stress; IMP.
DR InterPro; IPR003124; WH2.
DR Pfam; PF02205; WH2; 2.
DR SMART; SM00246; WH2; 2.
FT DOMAIN 5 14 POLY-PRO.
FT DOMAIN 239 245 POLY-PRO.
FT DOMAIN 349 357 POLY-PRO.
FT DOMAIN 396 406 POLY-PRO.
FT DOMAIN 424 431 POLY-PRO.
FT DOMAIN 462 468 POLY-SER.

```

```
FT DOMAIN 704 708 POLY-PRO.
FT CONFLICT 308 308 P -> R (IN REF. 1).
FT CONFLICT 350 350 A -> R (IN REF. 1).
FT CONFLICT 689 689 V -> E (IN REF. 1).
FT CONFLICT 710 817 PSTMDTSTNSPSKLNKORLFSTGGSTLOKHNTHTNQPDV
FT DVERVYTGGSNSIVGAKSGNERIVDDSPFKNTYSOMKPK
FT RPFONTKLYPSGKSGSVPLDILFT -> HUNIPVPLIA
FT FVKTLNNGYFLQVDRRCNTSIIRIQINQMLM (IN REF.
FT 1).
SQ SEQUENCE 817 AA; 82593 MW; 24C7522D5B1CA1C8 CRC64;

Alignment Scores:
Pred. No.: 5.73 Length: 817
Score: 83.50 Matches: 23
Percent Similarity: 48.57% Conservative: 11
Best Local Similarity: 32.86% Mismatches: 15
Query Match: 9.94% Indels: 21
DB: Gaps: 2

US-10-057-510-1 (1-486) x VRP1_YEAST (1-817)
Qy 294 CCAAGGCTGCATCTTCATCTCCAGCCCTCGAGCCTTCGTG----- 253
Db 406 ProglyAlaPheSerThrSerAlaLeuSerAlaSerileProLeuAlaProLeu 425
Qy 252 -----CCCTCTCCAGGCTCGCTCCATCTCCAGCCCTCGAGCCTTCCTCA 205
Db 426 ProProProProProSerValAlaThrSerValProSerAlaProProProPro 445
Qy 204 CACGTCATGCTTCCACCTTGCACATGGANAGCCGCTGCCCCACAGCCAGAGGGA 145
Db 446 ThrLeu-----ThrThrAsnLysProSerAla 454
Qy 144 ACACGAGCTGGCAAGAAAGACAGCAGT 115
Db 455 SerSerLysGlnSerLysLysSerSer 464

RESULT 6
ID STR5_HUMAN STANDARD; PRT; 213 AA.
AC Q9NSY2; P59094;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE STAR-related lipid transfer protein 5 (STARD5) (START domain-
DE containing protein 5).
GN STARD5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Liver;
EX MEDLINE=22008095; PubMed=12011452;
RA Soccio R.E., Adams R.M., Romanowski M.J., Sehayek E., Burley S.K.,
RA Breslow J.L.;
RT "The cholesterol-regulated StarD4 gene encodes a STAR-related lipid
RT transfer protein with two closely related homologues, Stard5 and
RT StarD6."
RL Proc. Natl. Acad. Sci. U.S.A. 99:6943-6948 (2002).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Testis;
RA Koehrer K., Beyer A., Mewes H.-W., Gassenhuber J., Wiemann S.;
RA Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Small intestine;
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RT "NEDO human cDNA sequencing project."
```

```
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Pancreas;
RX MEDLINE=22389257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haie F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Prange C.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Mullahy S.J.,
RA Raha S.S., Lequellano N.A., Peters G.J., Abramson R.D., Gunaratne P.H.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
CC -!- FUNCTION: May be involved in the intracellular transport of
CC sterols or other lipids. May bind cholesterol or other sterols (By
CC similarity).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC Name=2;
CC IsoId=Q9NSY2-1; Sequence=Displayed;
CC IsoId=Q9NSY2-2; Sequence=VSP_006278;
CC Note=No experimental confirmation available;
CC -!- SIMILARITY: Contains 1 START domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF480304; AAL89654.1; -
DR EMBL; AL137657; CAB70862.1; -
DR EMBL; AK026352; -; NOT ANNOTATED CDS.
DR EMBL; BC004365; AAH04365.1; ALT_INIT.
DR PIR; T46357; T46357 STARD5.
DR Genew; HGNC:18065; STARD5.
DR MIM; 607050; -
DR InterPro; IPR000799; STAR.
DR InterPro; IPR002913; STAR.
DR Pfam; PF01852; START; 1.
DR PRINTS; PR00978; STARPROTEIN.
DR SMART; SM02334; START; 1.
DR PROSITE; PS00448; START; 1.
KW Transposin; Lipid transport; Lipid-binding; Alternative splicing.
FT DOMAIN 1 213
FT VARSPLIC 1 107 Missing (in isoform 2).
FT SEQUENCE 213 AA; 23794 MW; EE5547A7846AB3C4 CRC64;

Alignment Scores:
Pred. No.: 5.75 Length: 213
Score: 83.00 Matches: 40
Percent Similarity: 33.55% Conservative: 12
Best Local Similarity: 25.81% Mismatches: 60
Query Match: 9.88% Indels: 43
DB: Gaps: 6

US-10-057-510-1 (1-486) x STR5_HUMAN (1-213)
```



QY 456 TGCNAAAACCA-----CGGGNCTFANGANCAGTTTTCAGACAG----- 415  
 Db CysValLysProAlaValGlyGlyLeuArgValLysTrpAspGluAsnValThrGlyPhe 86  
 QY 414 -----AGGGCCAGCACCAACACAGGAGGGG 388  
 Db GluIleIleGlnSerIleThrAspThrLeuCysValSerArgThrProSerAla 106  
 QY 387 GCTTCTTATGCAAGTTCCTTCATGCCATATGAGCCAAAGNCTNGTGGCTCGAACCCCTT 328  
 Db AlaMetLysLeuIleSerProArgAspPheValAspLeuValLeuValAspGlyTyrGlu 126  
 QY 327 GACTGGGCA-----TGTCTGAGAGTCT 304  
 Db Asp-GlyThrIleSerSerAsnAlaThrHisValGluHisProLeuCysProProLysPhe 146  
 QY 303 GGGGACNATCCAGGGCTCATCTTCATCTCAGCCCTCGGTAGCTTCTGCGCTCCCTCC 244  
 Db OglyPheValArgGlyPheAsn-----HisProCysGlyCysPheCysGluProLe 163  
 QY 243 CAGGGCTGCTGCATTC-----CTTCAGCTCAGGGCTCTCCACAGCTCCATGCCCTTCCAC 187  
 Db uProGlyGluProThrLysThrAsnLeuValThrPhePheHisThrAspLeuSerGlyTy 183  
 QY 186 CTTGCCACATGGANACAGCGCTGCCCCACAGCCAGGAGGGAACACGAAGCTGCCAGA 127  
 Db rLeuPro-----GlnAsnValValAspSerPhePhe 194  
 QY 126 AAGAACAGCAGTGTAGCAGCGGGCAACAATGAGAAAGGAGTA 84  
 Db aArgSerMetThrArgPheTyrAlaAsnLeuGlnLysAlaVal 208

RESULT 7  
 DUFF\_PANTR STANDARD; PRT; 336 AA.  
 AC Q95LF3;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Duffy antigen/chemokine receptor.  
 GN FY OR DARC.  
 OS Pan troglodytes (Chimpanzee).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.  
 OX NCBI\_TaxID=9598;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Tournamille C., Blancher A., Le Van Kim C., Gane P., Apoll P.A.,  
 RA Nakamoto W., Carlson J.-P., Colin Y.,  
 RT "Sequences, evolution and ligand binding properties of mammalian Duffy  
 RT antigen/receptor for chemokines."  
 RL Submitted (Oct-2000) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: Non-specific receptor for many chemokines (By  
 CC similarity).  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -!- SIMILARITY: Belongs to family Duffy of G-protein coupled  
 CC receptors.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AF311920; AAL09455.1;  
 DR InterPro; IPR005384; Duffyantigen.  
 DR PRINTS; PRO1559; DUFFYANTIGEN.  
 KW G-protein coupled receptor; Transmembrane; Glycoprotein.  
 FT DOMAIN 1 63 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 64 84  
 FT TRANSMEM 64 84 1 (POTENTIAL).

FT DOMAIN 85 95 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 96 116 2 (POTENTIAL).  
 FT DOMAIN 117 126 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 127 147 3 (POTENTIAL).  
 FT DOMAIN 148 166 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 167 187 4 (POTENTIAL).  
 FT DOMAIN 188 207 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 208 228 5 (POTENTIAL).  
 FT DOMAIN 229 244 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 245 265 6 (POTENTIAL).  
 FT DOMAIN 266 287 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 288 308 7 (POTENTIAL).  
 FT DOMAIN 309 336 CYTOPLASMIC (POTENTIAL).  
 FT CARBOHYD 16 16 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 27 27 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 33 33 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 336 AA; 35625 MW; 23067E1D8BF7CBF2 CRC64;

Alignment Scores:  
 Pred. No.: 5.95 Length: 336  
 Score: 83.00 Matches: 54  
 Percent Similarity: 37.23% Conservative: 16  
 Best Local Similarity: 28.72% Mismatches: 56  
 Query Match: 9.47% Indels: 62  
 DB: 1 Gaps: 11

US-10-057-510-1 (1-486) x DUFF\_PANTR (1-336)

QY 42 GGCTGGGAAGTGTGGTGTAAATGCCAAGCTCATATAGATACCTCT---TTCTCCAT 98  
 Db 95 GlyTrpProValLeu-----AlaGln-LeuAlaValGlySerAlaLeuPheSeril 111  
 QY 99 TGTGGCCCTGCTCTAACTGCTGCTCTTCTTCT-----TGCAGCT 137  
 Db 111 eValValProIleLeuAlaProGlyLeuGlySerThrArgSerSerAlaLeuCysSerLe 131  
 QY 138 TCCTTGTCTCTCTGCTT----- 156  
 Db 131 uGlyTyrCysValTrpTyrGlySerAlaPheAlaGlnAlaLeuLeuGlyCysHisAl 151  
 QY 157 -----GTGGGGCACGGCTGNTCCATGTGGCAAGGTGGAGGATGAGTGTGGAGGAG 212  
 Db 151 aSerLeuGlyHisArgLeuGly---AlaGlyGlnValProGlyLeu-ThrLeuGlyLeuT 170  
 QY 213 GCCTCGAGCTGAAGGA-----ATGCACGAGCCCTGGGAGGAGGGCAGAGGC 260  
 Db 170 hrValGlyIleTrpGlyValAlaAlaLeuLeu-ThrLeuProValThrLeuAlaSerGly 189  
 QY 261 TAGCAGGG-----CTGAGATGATGATGAGCCCTCGATNGTCC 302  
 Db 190 AlaSerGlyGlyLeuCysThrLeuIleTyrSerThrGluLeuLysAlaLeuGlnAlaThr 209  
 QY 303 CAGACTCTCAGGACATTGCCAGTCAAGGTTTCGAGCCACNAGNCTT---GGCTCATAT 359  
 Db 210 HisThrValAlaCysLeuAlaIlePheValLeuLeuProLeuGlyLeuPheGlyAlaLys 229  
 QY 360 GGCATGAGGGGAACATTGCATAGAAGCCCTCCCTCG-----TTGTGCTCTGG 410  
 Db 230 GlyLeuLysAlaLeuGlyMetGlyProGlyProTyrMetSerIleLeuTrpAlaTrp 249  
 QY 411 CCCTCTGTTCTGGAAACTGNTCTAGNCCCTCCCTGTTTTCGAAAGCCTCGAAGAG 470  
 Db 250 -PheIlePheTrp-----Trp----- 254  
 QY 471 GGAANTCCCTCGGG 486  
 Db 255 -----ProHisGly 257

RESULT 8  
 TLE2\_BRARE  
 ID TLE2\_BRARE STANDARD; PRT; 761 AA.  
 AC Q13166;  
 DT 15-JUL-1998 (Rel. 36, Created)



RP DEVELOPMENTAL STAGE.  
 RC TISSUE=Brain;  
 RX MEDLINE=99436166; PubMed=10506216;  
 RA Lim S., Naisbitt S., Yoon J., Hwang J.I., Suh P.G., Sheng M., Kim E.;  
 RT "Characterization of the Shank family of synaptic proteins. Multiple  
 RT genes, alternative splicing, and differential expression in brain and  
 RT development."; J. Biol. Chem. 274:29510-29518(1999).  
 RL J. Biol. Chem. 274:29510-29518(1999).  
 RN PARTIAL SEQUENCE FROM N.A. (ISOFORM 4).  
 RP TISSUE=Brain;  
 RX MEDLINE=20020275; PubMed=10551867;  
 RA Zitzer H., Hoenck H.-H., Baechner D., Richter D., Kreienkamp H.-J.;  
 RT "Somatostatin receptor interacting protein defines a novel family of  
 RT multidomain proteins present in human and rodent brain."; J.  
 RL J. Biol. Chem. 274:32997-33001(1999).  
 RN [6]  
 RP INTERACTION WITH HOMER-1, AND SUBCELLULAR LOCATION.  
 RX MEDLINE=99360651; PubMed=10433269;  
 RA Tu J.C., Xiao B., Naisbitt S., Yuan J.P., Petralia R.S., Brakeman P.,  
 RA Doan A., Asakura V.K., Lanahan A.A., Sheng M., Worley P.F.;  
 RT "Coupling of mGluR/Homer and PSD-95 complexes by the Shank family of  
 RT postsynaptic density proteins."; Neuron 23:583-592(1999).  
 RL Neuron 23:583-592(1999).  
 RN [7]  
 RP INTERACTION WITH SPTAN1.  
 RX MEDLINE=21523912; PubMed=11509555;  
 RA Bockers T.M., Mameza M.G., Kreutz M.R., Bockmann J., Weise C.,  
 RA Buck F., Richter D., Gundelfinger E.D., Kreienkamp H.-J.;  
 RT "Synaptic scaffolding proteins in rat brain. Ankryrin repeats of the  
 RT multidomain Shank protein family interact with the cytoskeletal  
 RT protein alpha-fodrin."; J. Biol. Chem. 276:40104-40112(2001).  
 RL J. Biol. Chem. 276:40104-40112(2001).  
 RN [8]  
 RP FUNCTION.  
 RX MEDLINE=21389514; PubMed=11498055;  
 RA Sala C., Piech V., Wilson N.R., Passafaro M., Liu G., Sheng M.;  
 RT "Regulation of dendritic spine morphology and synaptic function by  
 RT Shank and Homer."; Neuron 31:115-130(2001).  
 RL Neuron 31:115-130(2001).  
 RN [9]  
 RP REVIEW.  
 RX MEDLINE=20267867; PubMed=10806096;  
 RA Sheng M., Kim E.;  
 RT "The Shank family of scaffold proteins."; J.  
 RL J. Cell Sci. 113:1851-1856(2000).  
 CC -!- FUNCTION: Seems to be an adapter protein in the postsynaptic  
 CC density (PSD) of excitatory synapses that interconnects receptors  
 CC of the postsynaptic membrane including NMDA-type and metabotropic  
 CC glutamate receptors, and the actin-based cytoskeleton. May play a  
 CC role in the structural and functional organization of the  
 CC dendritic spine and synaptic junction. Overexpression promotes  
 CC maturation of dendritic spines and the enlargement of spine heads  
 CC via its ability to recruit Homer to postsynaptic sites, and  
 CC enhances presynaptic function.  
 CC -!- SUBUNIT: May homodimerize via its SAM domain. Interacts with  
 CC SPTAN1, Homer-1 and DLGAP1/GKAP. Is part of a complex with  
 CC DLG4/PSD-95 and DLGAP1/GKAP. Interacts with SSTR2 C-terminus via  
 CC the PDZ domain (by similarity).  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic; postsynaptic density of  
 CC neuronal cells. Colocalizes with alpha-latrotoxin receptor 1.  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=5;  
 CC Comment=Additional isoforms seem to exist;  
 CC Name=1;  
 CC IsoId=Q9WV48-1; Sequence=VSP\_006072; VSP\_006073;  
 CC Name=2;  
 CC IsoId=Q9WV48-2; Sequence=VSP\_006072; VSP\_006073;  
 CC Name=3;  
 CC IsoId=Q9WV48-3; Sequence=VSP\_006074;  
 CC Name=4; Synonyms=A;  
 CC IsoId=Q9WV48-4; Sequence=VSP\_006075;  
 CC Name=5;  
 CC IsoId=Q9WV48-5; Sequence=VSP\_006076; VSP\_006077;  
 CC TISSUE SPECIFICITY: Expressed only in brain (neuropil of cortex,  
 CC CA1 region hippocampus and molecular layer of cerebellum).  
 CC -!- DEVELOPMENTAL STAGE: Expression increases from low levels at birth  
 CC to high levels at 3-4 weeks before dropping slightly in adulthood.  
 CC Expressed in the cortex and the molecular layer of the cerebellum  
 CC at postnatal day 7. Isoform 2 expression does not change during  
 CC development of both cortex and cerebellum. Isoform 4 expression  
 CC decreases significantly during development of cortex but not  
 CC cerebellum.  
 CC -!- SIMILARITY: BELONGS TO THE SHANK FAMILY.  
 CC -!- SIMILARITY: Contains 7 ANK repeats.  
 CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.  
 CC -!- SIMILARITY: Contains 1 sterile alpha motif (SAM) domain.  
 CC -!- SIMILARITY: Contains 1 SH3 domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AF102855; AAD04569.2; .  
 CC EMBL; AF131951; AAD29417.1; ALT\_INIT.  
 CC EMBL; AF159046; AAD42975.1; .  
 CC EMBL; AF141904; AAF02498.1; ALT\_INIT.  
 CC HSSP; P00519; IABL.  
 CC InterPro; IPR002110; ANK.  
 CC InterPro; IPR001478; PDZ.  
 CC InterPro; IPR001660; SAM.  
 CC InterPro; IPR001452; SH3.  
 CC Pfam; PF00023; ank; 7.  
 CC Pfam; PF00595; PDZ; 1.  
 CC Pfam; PF00536; SAM; 1.  
 CC Pfam; PF00018; SH3; 1.  
 CC PRODOM; PD000066; SH3; 1.  
 CC SMART; SMO0248; ANK; 6.  
 CC SMART; SMO0228; PDZ; 1.  
 CC SMART; SMO0454; SAM; 1.  
 CC SMART; SMO0326; SH3; 1.  
 CC PROSITE; PS00088; ANK\_REPEAT; 3.  
 CC PROSITE; PS0297; ANK\_REPEAT\_REGION; 1.  
 CC PROSITE; PS0106; PDZ; 1.  
 CC PROSITE; PS0002; SH3; 1.  
 CC PROSITE; PS0105; SAM\_DOMAIN; 1.  
 CC KW ANK repeat; SH3 domain; Repeat; Alternative splicing.  
 CC FT REPEAT 195 210  
 CC FT REPEAT 212 245  
 CC FT REPEAT 246 278  
 CC FT REPEAT 279 312  
 CC FT REPEAT 313 345  
 CC FT REPEAT 346 378  
 CC FT REPEAT 379 395  
 CC FT DOMAIN 554 613  
 CC FT DOMAIN 663 757  
 CC FT DOMAIN 2104 2167  
 CC FT DOMAIN 929 932  
 CC FT DOMAIN 1010 1015  
 CC FT DOMAIN 1022 1027  
 CC FT DOMAIN 1194 1199  
 CC FT DOMAIN 1850 1860  
 CC FT VARSPLIC 1 613  
 CC FT VARSPLIC 615 654  
 CC FT VARSPLIC 646 654  
 CC FT VARSPLIC 797 804  
 CC FT VARSPLIC 797 804

CC IsoId=Q9WV48-5; Sequence=VSP\_006076, VSP\_006077;  
 CC TISSUE SPECIFICITY: Expressed only in brain (neuropil of cortex,  
 CC CA1 region hippocampus and molecular layer of cerebellum).  
 CC -!- DEVELOPMENTAL STAGE: Expression increases from low levels at birth  
 CC to high levels at 3-4 weeks before dropping slightly in adulthood.  
 CC Expressed in the cortex and the molecular layer of the cerebellum  
 CC at postnatal day 7. Isoform 2 expression does not change during  
 CC development of both cortex and cerebellum. Isoform 4 expression  
 CC decreases significantly during development of cortex but not  
 CC cerebellum.  
 CC -!- SIMILARITY: BELONGS TO THE SHANK FAMILY.  
 CC -!- SIMILARITY: Contains 7 ANK repeats.  
 CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.  
 CC -!- SIMILARITY: Contains 1 sterile alpha motif (SAM) domain.  
 CC -!- SIMILARITY: Contains 1 SH3 domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AF102855; AAD04569.2; .  
 CC EMBL; AF131951; AAD29417.1; ALT\_INIT.  
 CC EMBL; AF159046; AAD42975.1; .  
 CC EMBL; AF141904; AAF02498.1; ALT\_INIT.  
 CC HSSP; P00519; IABL.  
 CC InterPro; IPR002110; ANK.  
 CC InterPro; IPR001478; PDZ.  
 CC InterPro; IPR001660; SAM.  
 CC InterPro; IPR001452; SH3.  
 CC Pfam; PF00023; ank; 7.  
 CC Pfam; PF00595; PDZ; 1.  
 CC Pfam; PF00536; SAM; 1.  
 CC Pfam; PF00018; SH3; 1.  
 CC PRODOM; PD000066; SH3; 1.  
 CC SMART; SMO0248; ANK; 6.  
 CC SMART; SMO0228; PDZ; 1.  
 CC SMART; SMO0454; SAM; 1.  
 CC SMART; SMO0326; SH3; 1.  
 CC PROSITE; PS00088; ANK\_REPEAT; 3.  
 CC PROSITE; PS0297; ANK\_REPEAT\_REGION; 1.  
 CC PROSITE; PS0106; PDZ; 1.  
 CC PROSITE; PS0002; SH3; 1.  
 CC PROSITE; PS0105; SAM\_DOMAIN; 1.  
 CC KW ANK repeat; SH3 domain; Repeat; Alternative splicing.  
 CC FT REPEAT 195 210  
 CC FT REPEAT 212 245  
 CC FT REPEAT 246 278  
 CC FT REPEAT 279 312  
 CC FT REPEAT 313 345  
 CC FT REPEAT 346 378  
 CC FT REPEAT 379 395  
 CC FT DOMAIN 554 613  
 CC FT DOMAIN 663 757  
 CC FT DOMAIN 2104 2167  
 CC FT DOMAIN 929 932  
 CC FT DOMAIN 1010 1015  
 CC FT DOMAIN 1022 1027  
 CC FT DOMAIN 1194 1199  
 CC FT DOMAIN 1850 1860  
 CC FT VARSPLIC 1 613  
 CC FT VARSPLIC 615 654  
 CC FT VARSPLIC 646 654  
 CC FT VARSPLIC 797 804  
 CC FT VARSPLIC 797 804

CC SQGRQESRDKAKRLFRHVTGVSYSDFDAPSLIDGDSG  
 CC -> MALSAVGCGGGLPQFPFALSSSWPAGFRRSVWY  
 CC /FTid=VSP\_006073.  
 CC /FTid=VSP\_006073.  
 CC /FTid=VSP\_006074.  
 CC /FTid=VSP\_006075.  
 CC /FTid=VSP\_006075.

```

FT VARSPLIC 1930 1943 LSEDSQTSLLSKPS -> QYRIWVKSSDFGDF (in isoform 5)
FT /FTID=VSP 006076.
FT Missing (in isoform 5).
FT VARSPLIC 1944 2167 /FTID=VSP 006077.
FT S -> T (IN REF. 1).
FT CONFLICT 1141 1141 S -> N (IN REF. 2).
FT CONFLICT 1174 1174 S -> N (IN REF. 2).
FT CONFLICT 1246 1246 R -> K (IN REF. 1).
FT CONFLICT 1323 1323 A -> T (IN REF. 1).
FT CONFLICT 1331 1331 S -> D (IN REF. 1).
FT CONFLICT 1726 1726 S -> N (IN REF. 2).
SQ SEQUENCE 2167 AA; 226333 MW; 3F478B5A7B18BA86 CRC64;

Alignment Scores:
Pred. No.: 6 85 Length: 2167
Score: 83.00 Matches: 34
Percent Similarity: 36.79% Conservative: 5
Best Local Similarity: 32.08% Mismatches: 34
Query Match: 9.47% Indels: 33
DB: 1 Gaps: 5

US-10-057-510-1 (1-486) x SHK1_RAT (1-2167)
QY 195 CATGACGTGTGAGAGGCGCTGAGCTGAGAAATGACGAGCCCTGGGAGGAGGGA 254
Db H15AaAaHisGluProValLeuArgLeu-----TrpGlyAspProAlaArgGlu 1425
QY 255 GAAGGCTAC---CGAGGCTGAGGATGAGATGACGCCCTGGATNGT-----CCCCAG 305
Db LeuGlyTyArgAlaGlyLeuGlySerGlnGluLysAlaLeuThrAlaSerProAla 1445
QY 306 ACTCTCAGGACATTCGCCAGTCAAGGGTTGAGCCACNAGGNTTGGCTCATATGGCATG 365
Db AlaArgArgSerLeu-----
QY 1446 AlaArgArgSerLeu-----1450
QY 366 AAGCGGAACTGTGATAGAGAGCCCTCCCTGTTGCTGCTGCGCCCTCTGTTCTGGAA 425
Db -----LeuHisArgLeuProPro-ThrAlaProGlyValGlyProLeuLeuLeu 1467
QY 426 AACTGGTNTAGNCCCC-----TGCTTTTNGCAA 458
Db nLeuGlyProGluProProThrProHisProGlyValSerLysAlaTyrArgThrAlaAl 1487
QY 459 CCCTGAAGAAGGAA 474
Db aProGluGluProGlu 1492

RESULT 10
KCH2_RABIT STANDARD; PRT; 1161 AA.
AC Q8WNY2; 502731; 019119; 097586; Q9TV06;
DT 28-FEB-2003 (Rel. 41, Created)
DT 15-MAR-2004 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Potassium voltage-gated channel subfamily H member 2 (Ether-a-go-go related gene potassium channel 1) (ERG1) (RERG) (ra-erg) (Ether-a-go-go related protein 1) (Eag related protein 1).
DE KCHN2 OR ERG.
GN Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OK NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RA Wischel H.J., Hancock J.C., Levi A.J., Meech R.W.;
RT "RERG - rabbit ventricular ERG K+ channel subunit.";
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 411-571 FROM N.A. (ISOFORM 1/2).
RA MEDLINE=97164986; PubMed=9012748;
RP Wymore R.S., Gintant G.A., Wymore R.T., Dixon J.E., McKinnon D.,
RA Cohen I.S.;
RT "Tissue and species distribution of mRNA for the IKr-like K+ channel,"

```

```

RT
RL
CC CIRC. Res. 80:261-268(1997).
CC FUNCTION: Pore-forming (alpha) subunit of voltage-gated inwardly rectifying potassium channel. Channel properties are modulated by cAMP and subunit assembly. Mediates the rapidly activating component of the delayed rectifying potassium current in heart (IKr) (By similarity).
CC SUBUNIT: The potassium channel is probably composed of a homo- or heterotetrameric complex of pore-forming alpha subunits that can associate with modulating beta subunits. Heteromultimer with KCHN6/ERG2, KCHN7/ERG3, KCNE1 and KCNE2 (By similarity).
CC SUBCELLULAR LOCATION: Integral membrane protein.
CC ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q8WNY2-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q8WNY2-2; Sequence=VSP_000971;
CC TISSUE SPECIFICITY: Detected in heart, both in atrium and in left ventricle.
CC DOMAIN: The segment S4 is probably the voltage-sensor and is characterized by a series of positively charged amino acids at every third position.
CC PTM: Phosphorylated on serine and threonine residues (By similarity).
CC SIMILARITY: Belongs to the potassium channel family. H (Eag) subfamily.
CC SIMILARITY: Contains 1 cyclic nucleotide-binding domain.
CC SIMILARITY: Contains 1 PAS (PER-ARNT-SIM) dimerization domain.
CC SIMILARITY: Contains 1 PAS-associated C-terminal (PAC) domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC
CC EMBL; U87513; AAB68612.1; -.
CC EMBL; AF068736; AAC99425.1; -.
CC EMBL; AF105061; AAD39357.1; -.
CC EMBL; U75212; AAC48723.1; -.
CC InterPro; IPR000595; CNMP binding.
CC InterPro; IPR003967; Erg_Channel.
CC InterPro; IPR005821; Ion_Channel.
CC InterPro; IPR001622; K_Channel_pore.
CC InterPro; IPR005820; M_Channel_nlg.
CC InterPro; IPR001610; PAC.
CC InterPro; IPR000700; PAS-associ C.
CC InterPro; IPR000014; PAS domain.
CC Pfam; PF00027; CNMP_binding; 1.
CC Pfam; PF00520; Ion_Trans; 1.
CC Pfam; PF00785; PAC; 1.
CC PRINTS; PR01470; ERGCHANNEL.
CC SMART; SM00100; CNMP; 1.
CC SMART; SM00086; PAC; 1.
CC PROSITE; PS00888; CNMP_BINDING_1; FALSE NEG.
CC PROSITE; PS00889; CNMP_BINDING_2; FALSE NEG.
CC PROSITE; PS00042; CNMP_BINDING_3; 1.
CC PROSITE; PS01112; PAS; 1.
CC PROSITE; PS01113; PAC; 1.
CC Transport; Ion transport; 1.
CC Potassium channel; Potassium; Voltage-gated channel;
KW Phosphorylation; Glycoprotein; Multigene family; Alternative splicing.
FT DOMAIN 1 405 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 406 426 SEGMENT S1 (POTENTIAL).
FT TRANSMEM 453 473 SEGMENT S2 (POTENTIAL).
FT DOMAIN 474 497 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 498 518 SEGMENT S3 (POTENTIAL).
FT TRANSMEM 523 543 SEGMENT S4 (POTENTIAL).
FT DOMAIN 544 549 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 550 570 SEGMENT S5 (POTENTIAL).

```



```

Percent Similarity: 38.46%      Conservative: 25
Best Local Similarity: 20.98%    Mismatches: 55
Query Match: 9.82%              Indels: 33
DB: 1                           Gaps: 4

US-10-057-510-1 (1-486) x SPRA_CABEL (1-1311)
QY 423 CAGAGAGAGGGCCAGGACACACACAGGAGGGGCTTCT-----382
Db 921 ProGluLysArgLysGlnSerSerGluGluProAlaSerAspProGluLeuPheGly 940
QY 381 TATCAGAGTCCCTTCATGCCATATGAGCAGCAAGNCTGTGGCTCGAACCTTGACTGG 322
Db 941 SerSerSerGlnProThrArgGlnLeuSerGluArgAlaThrArgAsnArgIleAsnTyr 960
QY 321 GCAATGTCCTAGAGTCTGGGACNATCCAGGGCTGCATCTTCATCCTCAGCGCTGCGT 262
Db 961 SerLeuLeuSerLysAsnGlySerGlyLysProThrProSerThrSerSerAlaAsnLeu 980
QY 261 AGCCTTCTGCCCTCCAGGCTCTCCATTCCTTCAGCTCCAGGCGCTCTCCACAC 202
Db 981 GluLysLeuAlaGlySerSerGlyLysSerGluSerProGluProAspGluSer 1000
QY 201 GTCCATGCTCTCCAC-----CTGCCACATGGANACAG-----169
Db 1001 ValGluValSerHisTrpLysIleArgThrPheLeuArgSerGluTyrGlyValLysGlu 1020
QY 168 -----CCGTGCCCCACAGCCAGGAGGACACAGAGCTGGCAAGAG 124
Db 1021 SerLeuLysCysProAspCysProTyrLysSerSerGluProAspValLeu-----1037
QY 123 AACAGCAGTTAGACGAGGGGCAACATGGAGAGGAGTATCTATTATGAGCTTGGGCAT 64
Db 1038 -----GluLysHisArgTyrTyrHisMetThrLys 1047

RESULT 12
DUFF_PAPHA
ID DUFF_PAPHA STANDARD; PRT; 336 AA.
AC Q95LIG5;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Duffy antigen/chemokine receptor.
GN FY OR DARC
OS Papio hamadryas (Hamadryas baboon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Papio.
OX NCBI_TaxID=9557;
RN [1]
RP SEQUENCE FROM N.A.
RA Tournamille C., Blancher A., Le Van Kim C., Gane P., Apoil P.A.,
RA Nakamoto W., Carlson J.-P., Colin Y.;
RT "Sequences, evolution and ligand binding properties of mammalian duffy
RT antigen/receptor for chemokines";
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Non-specific receptor for many chemokines (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: Belongs to family Duffy of G-protein coupled
CC receptors.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).

```

```

CC EMBL; AF303532; AAL09324.1; -
DR InterPro; IPR005384; DuffyAntigen.
DR PRINTS; PR01559; DUFFYANTIGEN.
KW G-protein coupled receptor; Transmembrane; Glycoprotein.
FT DOMAIN 1 63 1 (POTENTIAL).
FT TRANSMEM 64 84 1 (POTENTIAL).
FT DOMAIN 85 95 2 (POTENTIAL).
FT TRANSMEM 96 116 2 (POTENTIAL).
FT DOMAIN 117 126 3 (POTENTIAL).
FT TRANSMEM 127 147 3 (POTENTIAL).
FT DOMAIN 148 166 4 (POTENTIAL).
FT TRANSMEM 167 187 4 (POTENTIAL).
FT DOMAIN 188 207 5 (POTENTIAL).
FT TRANSMEM 208 228 5 (POTENTIAL).
FT DOMAIN 229 244 6 (POTENTIAL).
FT TRANSMEM 245 265 6 (POTENTIAL).
FT DOMAIN 266 287 7 (POTENTIAL).
FT TRANSMEM 288 308 7 (POTENTIAL).
FT DOMAIN 309 336 7 (POTENTIAL).
FT CARBOHYD 16 16 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 27 27 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 33 33 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 336 AA; 35603 MW; 5B85DAD812DB685 CRC64;

Alignment Scores:
Pred. No.: 7.35      Length: 336
Score: 82.00      Matches: 55
Percent Similarity: 37.97%      Conservative: 16
Best Local Similarity: 29.41%      Mismatches: 56
Query Match: 9.36%      Indels: 60
DB: 1      Gaps: 12

US-10-057-510-1 (1-486) x DUFF_PAPHA (1-336)
QY 42 GCGTGGAGTGTGTGCTGTAATAATGCCCAAGCTCATATAGATCTCTCTCTCCAT 98
Db 95 GlyTrpProValLeu-----AlaGln-LeuAlaValGlySerAlaLeuPheSerI 111
QY 99 TGTGCCCCCTGCTTAACGTGCTTCTTCT-----TGCAGCT 137
Db 111 eValValProIleLeuAlaProGlyLeuGlyAsnThrArgSerSerAlaLeuCysSerLe 131
QY 138 TCGTGTTCCTCTCTG-----CTGTGGGG-----GC 164
Db 131 uGlyTyrCysValTrpTyrGlySerAlaPheAlaGlnAlaLeuLeuGlyCysHisAl 151
QY 165 ACGGCTGTNTCCA-----TGTGCAAGGTGGAGGCGATGCGTGTGCGAGGCGC 215
Db 151 aserLeuGlyProLysLeuGlyAlaAspGlnValProGlyLeu-ThrLeuGlyLeuServ 171
QY 216 CTGGAGCTGAAGGA-----ATGGACGAGCCCTGGGAGGAGGCGAGAGGCTAC 263
Db 171 alGlyLeuTrpGlyValAlaAlaLeuLeu-ThrLeuProValThrLeuAlaSerGlyAla 190
QY 264 GCAGG-----CTGAGGATGAAGATGCAGCCCTCGATNGTCCCCAG 305
Db 191 SerGlyGlyLeuCysThrProValTyrSerMetGluLeuLysAlaLeuGlnAlaThrHis 210
QY 306 ACTCTCAGGACATTCGCCAGCTCAAGGTTTCGAGGACACNAGNCTT---GGCTCATATGC 362
Db 211 AlaValAlaCysLeuAlaIlePheValLeuLeuProLeuGlyLeuPheGlyAlaLysGly 230
QY 363 ATGAAGGGGAACCTTCATAGAAGCCCTCCCTGG-----TTGTGTCTCTGGGCC 413
Db 231 LeuLysLysAlaLeuGlyMetGlyProGlyProIleMetAsnIleLeuTrpAlaTrp-P 250
QY 414 TCTGTCTCGAAAACACTGGTNTAGNCCCTCGTTTNTGCAAGACCTGAAGAAGGGA 473
Db 250 eilePheTrp-----Trp-----254
QY 474 AATCCCCCATGGG 486

```







Search completed: April 6, 2004, 17:56:04  
Job time : 20 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: April 6, 2004, 17:52:30 ; Search time 17 Seconds

(without alignments)  
5499.892 Million cell updates/sec

Title: US-10-057-510-1

Perfect score: 876

Sequence: 1 gcaacagcgatnaaaaaat.....aaagggaantccccatggg 485

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 566732

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.spool -DEV=xlp  
-Q=/cgn2.1/USPTO.spool\_p/US10057510/runat\_06042004\_141609\_14737/app.query.fasta\_1.647  
-DB=PIR 78 -QFMT=fastan -SUFFIX=rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10057510@cgn 1.1.38 @runat\_06042004\_141609\_14737 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DSV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR 78.\*

1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	84.5	10.1	1173	2 T31421	C-terminal domain-
C 2	84.5	10.1	1355	2 S40022	spalt protein - fr
C 3	84.5	10.1	2957	2 T33152	hypothetical prote
C 4	84	9.6	338	2 I52608	glycoprotein D - h
C 5	83.5	9.9	383	2 T46707	proteophosphoglyca
C 6	83.5	9.9	817	2 S51342	verprolin - yeast
C 7	83	9.9	1706	2 T56333	probable RNA helic
C 8	82.5	9.8	325	2 T96815	hypothetical prote
C 9	82.5	9.8	749	2 H08101	outer dynein arm d
C 10	82.5	9.8	1309	2 T19170	hypothetical prote
C 11	81	9.6	309	1 S03257	homeotic protein H
C 12	81	9.2	410	2 AG3968	hypothetical prote
C 13	81	9.2	410	2 D98314	hypothetical prote
C 14	81	9.6	772	2 T13078	KIAA0992 protein -

C 15	81	9.6	1791	2 T02345	hypothetical prote
C 16	80.5	9.6	217	2 B41256	sdr protein - chic
C 17	80.5	9.6	392	2 B48423	homeotic protein e
C 18	80	9.5	328	2 B88979	protein F37B4.10 l
C 19	80	9.5	360	1 A25732	inhibin alpha chai
C 20	80	9.5	540	1 F0MVHL	gag polyprotein -
C 21	79.5	9.5	268	2 A49303	homeotic protein C
C 22	79.5	9.5	401	2 A48423	engrailed homeodom
C 23	79.5	9.5	450	2 S00950	hypothetical prote
C 24	79.5	9.5	1239	2 G71266	probable ATP-depen
C 25	79.5	9.5	1309	2 T00078	probable RNA-direc
C 26	79	9.0	340	2 C69328	iron-sulfur cluste
C 27	78.5	9.0	107	2 T35634	hypothetical prote
C 28	78.5	9.3	178	2 G84357	probable acetyltra
C 29	78.5	9.3	1026	1 A40315	maternal effect pr
C 30	78.5	9.3	1734	2 A54602	microtubule-associ
C 31	78	8.9	154	2 T03956	hypothetical prote
C 32	78	9.3	834	2 S19155	serotonin receptor
C 33	77.5	9.2	173	2 A72450	hypothetical prote
C 34	77.5	8.8	865	2 T40170	hypothetical prote
C 35	77.5	9.2	1446	1 A45344	immediate-early pr
C 36	77	9.2	1060	2 S33641	homeotic protein z
C 37	76.5	8.7	275	2 T50378	hypothetical prote
C 38	76.5	9.1	301	2 T03914	hypothetical prote
C 39	76.5	8.7	652	2 D65215	acetate-CoA ligase
C 40	76.5	8.7	652	2 C91260	acetyl-CoA synthet
C 41	76.5	8.7	652	2 G86100	acetyl-CoA synthet
C 42	76.5	9.1	1017	2 T48452	hypothetical prote
C 43	76	9.0	654	2 B43553	membrane protein x
C 44	76	9.0	1032	2 T34433	hypothetical prote
C 45	76	9.0	2232	2 T34434	hypothetical prote

ALIGNMENTS

RESULT 1

T31421 C-terminal domain-binding protein rA1 - rat

C;Species: Rattus norvegicus (Norway rat)

C;Date: 29-Oct-1999 #sequence\_revision 23-Oct-1999 #text\_change 07-Dec-1999

C;Accession: T31421

R;Yuryev, A.; Patturajan, M.; Litingtung, Y.; Joshi, R.V.; Gentile, C.; Gebata, M.; Cord

Proc.Natl. Acad. Sci. U.S.A. 93: 6975-6980, 1996

A;Title: The C-terminal domain of the largest subunit of RNA polymerase II interacts wit

A;Reference number: Z21024; MUID:96293459; PMID:8692929

A;Accession: T31421

A;Status: preliminary; translated from GE/EMBL/DBDJ

A;Residues: 1-1173 <YUR>

A;Cross-references: EMBL:U49056; NID:g1438531; PID:g1438532; PIDN:AAC52657.1

A;Experimental source: hippocampus

Alignment Scores:				
Pred. No.:	3.5	Length:	1173	
Score:	84.50	Matches:	54	
Percent Similarity:	39.43%	Conservative:	15	
Best Local Similarity:	30.86%	Mismatches:	47	
Query Match:	10.06%	Indels:	59	
DB:	2	Gaps:	12	

US-10-057-510-1 (1-486) x T31421 (1-1173)

QY	423	CCAGAACAGAGGGCCAGGACACACA-----ACCAGGAGGGGGGCTTCTTATGCAAGTTC	370
DB	65	ProArgLeuArgAlaTrpArgThrGlyThrValSerProGlnSerHisAlaSerArg	84
QY	369	CTTTCATGCATATGACCAAGNCCTTGCTGCAACCCCTTGACTGGGCAATGCTCTGA	310
DB	85	ProAlaCys-----SerArgHisLeuLeu-----ThrLeuGly-----	95
QY	309	GAGTCTGGGACNAT-----CCAGGGGCTGCATCTTCATCCTCAGCCCTCGGT	262
DB	96	---ThrGlyAspGlyGlyProAlaProProAlaProSerSerGlySerSerPro	114

```
QY 261 AGCTTCTGCTCCCTCCAGGCTGCTCCATTCCTTCAGCTCCAGCGCTCTCCACAC 202
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 115 SerProSerProSerSerSerProProProProProProProProProProProP 134
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 201 GTCCATGCTTCCA-----CCTGCCACATGG 175
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 134 roProAlaLeuProAlaProArgPheAspIleTyrAspProPheHisProThrAspGluA 154
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 174 ANACAGCGGTGCCCA-----CAAGCCAGAGGGAACA 142
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 154 laTyrSerProProAlaProGluGlnLysTyrAspProPheGluAlaThrGlySerA 174
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 141 AC-----GAAGCTGCCAAGAAGAAGACAGC 118
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 174 snProSerSerGlyThrProSerProGluGluGluGluGluGluGluG 194
QY 117 AGTTAGACAGGG-----CAACAATGGAGAAG-----GAGTATCTATTATGAGCTT 70
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 194 luGluGluGluGlyLeuSerGlnSerIleArgArgIleSerGluThrLeuAla----- 211
QY 69 GGCATTTTAC-----ACAGCACTTCCAGCTTTTCCA 35
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 212 --GlyIleTyrAspAsnSerLeuSerGlnAspPhePro 224
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
RESULT 2
S40022
spalt protein - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C>Date: 19-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 21-Jul-2000
C:Accession: S40022
R:Kuehlein, R.P.; Frommer, G.; Friedrich, M.; Gonzalez-Gaitan, M.; Weber, A.; Wagner-Ber
EMBO J. 13, 168-179, 1994
A>Title: spalt encodes an evolutionarily conserved zinc finger protein of novel structure
A:Reference number: S40022; MUID:94139659; PMID:7905822
A:Accession: S40022
A:Molecule type: DNA
A:Residues: 1-1355 <RUE>
A:Cross-references: EMBL:X75541; NID:9414105; PIDN:CAA53229.1; PID:92598394
A>Note: the authors translated the codon GAC for residue 51 as Ala
A>Note: mRNA sequencing has also been done
C:Genetics:
A:Gene: sal; spalt
A:Cross-references: FlyBase:FBgn0004579
A:Introns: 51/1; 1329/3 1355/2
C:Keywords: zinc finger
Alignment Scores:
Pred. No.: 3.51 Length: 1355
Score: 84.50 Matches: 46
Percent Similarity: 39.29% Conservative: 20
Best Local Similarity: 27.38% Mismatches: 66
Query Match: 10.06% Indels: 36
DB: 2 Gaps: 7
US-10-057-510-1 (1-486) x S40022 (1-1355)
QY 441 GGNCTANGANCAGTTTCCAGAACACAGAGGCCAGGACCACACACACAGGAGGGGGTCTT 382
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1029 GlyValValAsnThrAsnProValArgProArgSerSerAlaSerSerHisGlyHisSer 1048
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 381 TATGCAAGT-----TCCCTTCATGCTCATGATGCAAGNCTGNGGCT 337
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1049 ValGlySerThrSerAlaProThrSerProSerValHisAlaSerSerGlnValIleLys 1069
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 336 CGAACCCCTTGACTGGGCAATGTCTCAGAGTCTGGGGAC-----NATCCAGGG 289
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1069 ArgSerSerProAlaArgSerGluAlaSerGlnGlyAlaLeuAspLeuThrProArg 1088
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 288 GTTGCA---TCTTCATCTCAGCCCTCGTAGCTTCTGCCCTCTCCAGGGCTCGTCC 232
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1089 AlaAlaProThrSerSerSerSerArgSerProLeuProLysGluLysProValSer 1108
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
QY 231 ATTCCTTCAGCTCCAGCGC-----CTCTCCA 205
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1109 ProProSerLeuPro-ArgSerProSerGlySerSerHisAlaSerAlaAsnIleLeuTh 1128
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 204 CAGCTCCATGCTTCCACC-----TTGCCACATGGANACAGCGGTGCCCC 160
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1128 rSerProLeuProProThrValGlyIleAspCysLeuProLysGlyLeuGlnHis---Hi 1147
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 159 CACAAGCCAGAGGAACAACAGAGCTGGCAAGAAAGAACAGCAGCTTAGACAGGGGCAAC 100
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1147 sLeuGlnGlnGlnHisGlnHisLeuMetGlnGlnGlnAlaValaAlaAlaAlaA 1167
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 99 AATCGAAGAGGAGTATCTATTATGAGCTTGGCATTTTACCAGCACATTCCTCCAGCCTT 40
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1167 aAlaGlnHisHis-----HisGlnGlnMetAlaAlaLe 1179
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 39 TTCATAAACATCGAAACATT 18
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1179 uAspGlnHisGlnGlnLeu 1186
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
RESULT 3
T33152
Hypothetical protein T04D1.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T33152
R:Davidson, S.; Wohldmann, P.
Submitted to the EMBL Data Library, May 1998
A:Description: The sequence of C. elegans cosmid T04D1.
A:Reference number: Z21292
A:Accession: T33152
A>Status: preliminary; translated from GE/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2957 <DAV>
A:Cross-references: EMBL:AF067617; PIDN:AAC17559.1; GSPDB:GN00019; CESP:T04D1.4
A:Experimental source: strain Bristol N2; Clone T04D1
C:Genetics:
A:Gene: CESP:T04D1.4
A:Map position: 1
A:Introns: 122/3; 293/3; 515/3; 1205/2; 1577/1; 2221/3; 2776/1; 2864/3
Alignment Scores:
Pred. No.: 3.53 Length: 2957
Score: 84.50 Matches: 39
Percent Similarity: 40.44% Conservative: 16
Best Local Similarity: 28.68% Mismatches: 52
Query Match: 10.06% Indels: 30
DB: 2 Gaps: 5
US-10-057-510-1 (1-486) x T33152 (1-2957)
QY 420 GAACAGAGGGCCAGGACACACACAGGAGGGGCTTCTTATGCAAGTTCCTTCATGC 361
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2377 GlnGlnAlaGlnProSerThrSerLysGlyAlaSer---SerSerGlnSerSerAla 2395
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 360 CATATGAGCCAGNCTNGTGGCTCGAACCTTCTGCTGGGCAATG----- 316
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2396 GlnAlaAsnGlnAlaAlaGlnAlaGlnAlaAlaValaLeuAlaGlnValaAla 2415
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 315 -----TCTCGAGAGTCTGGGACNATCCAGGGGCTGCATCTTCATCTCCAGCC 268
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2416 AlaAlaSerAlaSerThrSerSerAlaAlaGlnAlaValaGlnGlnGlnVal 2435
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 267 CTGGTAGCTTCTG----- 253
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2436 LeuGluAlaLeuLeuAlaMetSerMetAsnProAlaAlaLeuAlaLeuThrGlyThr 2455
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 252 CCCTCTCTCCAGGGCTGCTCCATTCCTTCAGCTCCA-----GGCCTCTCTCCACAC 202
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2456 SerSerSerThrGlyThrSerThrProSerAlaProLysSerLysProAlaProGln 2475
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 201 GTCCATGCTTCCAGCTTGCACATG---GANACAGCGGTGCCCCCAAGCCAGAGGG 146
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

QY	42	GGCTGGGAAGTCTCGTGTAAATCCCAAGCTCATATAGATACTCCT---	TTCTCCAT	98
Db	97	GlyTrpProValLeu-----AlaGln-LeuAlaValGlySerAlaLeuPheSerI	113	
QY	99	TGTTGCCCTCGCTCTAACTGCTGTTCTTCT-----TGCCAGCT	137	
Db	113	evAlValProValLeuAlaProGlyLeuGlySerThrArgSerSerAlaLeuCysSerLe	133	
QY	138	TGCTTGTTCCTCCCTCGCTT-----	156	
Db	133	uGlyTyrcysValTrpTyrglySerAlaPheAlaGlnAlaLeuLeuGlyCysHisAl	153	
QY	157	-----GTGGGGGACGGCTGTNTCCATGTGGCAAGTGGGAAGCATGGACGTGTGGAGGAG	212	
Db	153	asrLeuGlyHisArgLeuGly--AlaGlyGlnValProGlyLeu-ThrLeuGlyLeu	172	
QY	213	CGCTGGAGCTGAAGGA-----ATGCAGACGCCCTGGGAGGAGGGCAGAGGC	260	
Db	172	hrValGlyLeTrpGlyValAlaAlaLeuLeu-ThrLeuProValThrLeuAlaSerGly	191	
QY	261	TACGCAGGG-----CTGAGGATGAAGATGCAGCCCTGGATGTCCTCC	302	

Qy	447	CGAGGGGNGTANGANCAGTTTCCTCCAGAACAGAGGCCACGACCACCAACACGAGGGAGGGG	388
Dd	11	ProSerAlaSerSerSerAlaPro-----SerSerSerSerAlaProSer	27
Qy	387	GCTTCTATGCAAGTTCCCTTCATGCATATGAGCAGCAAGNCCTGTGGCTTCGAACCCCTT	328
Dd	28	AlaSerSerSerSerAlaProSer-----SerSerSerAlaProSerAlaSerSer	45
Qy	327	GACTGGGCAATGCTCTGAGAGCTCTGGGACNATCCAGGGCTGCATTTCATCC-----	274
Dd	46	SerSerAlaProSerSerSerSerSerSerAlaProSerAlaSerSerSerAlaPro	65
Qy	273	TGACGCCCTCGTAGCTTCTGCCCTCTCTCCAGGGCTCGTCCCATTCCTTCAGCTCCAGCG	214
Dd	66	SerSerSerSerSerSerAlaProSerAlaSerSerSerSerAlaProSerSerSer	85
Qy	213	CCTCTTCACACAGTCCATGCTTCACCTTGCCACATGGANACAGCGTGCCCCCAAG	154
Dd	86	SerAlaPro-SerAlaSerSerSerAlaProSerSerSerSerAlaProSerAl	105
Qy	153	CCAGAGGGAAACAACGAGCTGGCAAGAAAAGAACAGCAGTCTAGACGAGGGGCAACA	99
Dd	105	aSerSerSerAlaProSerSerSerSerSerSerSerSerAlaProSerAlaSerSer	123

## RESULT 6

S51342  
verprolin - yeast (Saccharomyces cerevisiae)  
N:Alternate names: prolin-rich protein VRP1; protein L8300.13; protein YLR337c  
C:Species: Saccharomyces cerevisiae  
C:Date: 23-Feb-1995 #sequence revision 11-Aug-1995 #text\_change 23-Mar-2001  
C:Accession: S51342; S39626; S57435  
R:DNA, Z.  
submitted to the EMBL Data Library, December 1994  
A:Description: The sequence of *S. cerevisiae* cosmid 8300.  
A:Reference number: S51339  
A:Accession: S51342  
A:Molecule type: DNA  
A:Residues: 1-817 <DUZ>  
A:Cross-references: EMBL:U9028; NID:G609380; PID:G609392; MIPS:YLR337c  
R:Donnelly, S.F.H.; Focklington, M.J.; Pallotta, D.; Orr, E.  
Mol. Microbiol. 10, 585-596, 1993  
A:Title: A proline-rich protein, verprolin, involved in cytoskeletal organization and cell division in *S. cerevisiae*  
A:Reference number: S39626; MUID:95058201; PMID:7968536  
A:Accession: S39626  
A:Status: nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1-307, 'R', 309-349, 'R', 351-688, 'B', 690-709, 'HURWIPVPLIAPVKTLNNGYFLQVDRRCNWTSLI'  
A:Cross-references: EMBL:Z26645; NID:G414785; PIDN:CAA81388.1; PID:G439289  
R:Munn, A.L.; Stevenson, B.J.; Geli, M.I.; Riezman, H.  
submitted to the EMBL Data Library, June 1995  
A:Description: ends, end6, and end7: mutations that cause actin delocalization and block cell division  
A:Reference number: S57435  
A:Accession: S57435  
A:Molecule type: DNA  
A:Residues: 1-162, 'F', 164-817 <MUN>  
A:Cross-references: EMBL:X87806; NID:G871534; PID:G871535  
A:Experimental source: strain W303  
C:Genetics:  
A:Gene: SGD:VRP1; MDP5; END5  
A:Cross-references: SGD:S0004329; MIPS:YLR337c  
A:Map position: 12R  
F:5-14/Region: proline-rich  
F:77-85/Region: proline-rich  
F:114-180/Region: proline-rich  
F:216-245/Region: proline-rich  
F:305-336/Region: proline-rich  
F:349-357/Region: proline-rich  
F:372-382/Region: proline-rich  
F:396-406/Region: proline-rich  
F:421-445/Region: proline-rich  
F:518-528/Region: proline-rich  
F:567-577/Region: proline-rich  
F:608-621/Region: proline-rich  
F:649-661/Region: proline-rich  
F:678-685/Region: proline-rich  
F:704-710/Region: proline-rich  
Alignment Scores:  
Pred. No.: 4.39 Length: 817  
Score: 83.50 Matches: 23  
Percent Similarity: 48.57% Conservative: 11  
Best Local Similarity: 32.86% Mismatches: 15  
Query Match: 9.94% Indels: 21  
DB: 2 Gaps: 2  
US-10-057-510-1 (1-486) x S51342 (1-817)  
QY 294 CCAGGGCTCATCTTCATCTCAGCCCTCGTAGCTTCG----- 253  
Db 406 ProGlyAlaPheSerThrSerAlaLeuSerAlaSerSerIleProLeuAlaProLeu 425  
QY 252 -----CCCTCTCCAGGGCTCGTCCATTCCTTCAGCTCAGCGCTCTCCCA 205  
Db 426 ProProProProProProSerValAlaThrSerValProSerAlaProProProPro 445  
QY 204 CAGTCCTATCCCTTCACCTTGCCACATGGANACAGCGGTGCCCCCAAGCCAGAGGA 145

Db 446 ThrLeu-----ThrThrAsnLysProSerAla 454  
QY 144 ACAACAGAGCTGGCAAGAAAGAACAGCAGT 115  
Db 455 SerSerLysGlnSerLysValleSerSer 464  
RESULT 7  
B75633  
probable RNA helicase - *Deinococcus radiodurans* (strain R1)  
C:Species: *Deinococcus radiodurans*  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 31-Mar-2000  
C:Accession: B75633  
R:White, O.; Eisen, J.A.; Heideberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; M.  
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999  
A:Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1.  
A:Reference number: A75250; MUID:20036896; PMID:10567266  
A:Accession: B75633  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-1706 <WHI>  
A:Cross-references: GB:A8001826; NID:G6460827; PIDN:AAF12564.1; PID:G6460860; TIGR:DRB0.  
A:Experimental source: strain R1  
C:Genetics:  
A:Gene: DRB0135  
A:Map position: megaplasmid  
A:Genome: plasmid  
A>Note: plasmid MPI  
Alignment Scores:  
Pred. No.: 4.95 Length: 1706  
Score: 83.00 Matches: 40  
Percent Similarity: 33.52% Conservative: 20  
Best Local Similarity: 22.35% Mismatches: 52  
Query Match: 9.88% Indels: 67  
DB: 2 Gaps: 6  
US-10-057-510-1 (1-486) x B75633 (1-1706)  
QY 429 AGTTTCCAGAACAGAGGGCCAGGACACACACAGGAGGGGGCTTCTATGCAAGTTCC 370  
Db 818 ThrLeuProArgAspGluGlyArgSerThrGlyArgLeuSerLeuAlaGlnGlyAla 837  
QY 369 CCTTCATGCCATATGAGCCAAAGCCTGTGGCTCGAACCTTGACTGGGCAATCTCTGA 310  
Db 838 ValGlyArgTyrLeuThrArg-----ProgluThrLeuAlaTrpSerGlnAlaLeu 854  
QY 309 GAGTCTGGGAC----- 298  
Db 855 LysArgAspValAlaAlaLeuHisAspLeuLeuAlaLeuSerGluPheIle 874  
QY 297 -----NAT 295  
Db 875 GluGlnValGluProGlyAlaTyrGlnLeuLysGlyThrSerPheThrTrpHisAlaGly 894  
QY 294 CCAGGGCTGATCTTCATCTCAGCCCTCGTAGCTTCTGCGCTCTCCACGGGCTCG 235  
Db 895 ProGlyThrGluProSerValAlaLeuArgValLeuArgPro----- 909  
QY 234 TCCATTCCTTCAGCTCCAGCGCTCTCCACACCTCCATGCTTCCACCTTCCACATGG 175  
Db 910 -----AlaGlySerGluAlaProArgValAsnSerPheLeuSerLeuTyr 925  
QY 174 ANACAGCGTCCCGC-----CACAGCCAGAGGAAACACGAAGCTG--- 133  
Db 926 GlnGlnProAlaAlaGluPheArgAspLeuHisGlyAlaGluHisThrAlaGlnIleArg 945  
QY 132 -----GCAAGAAAGAACAGCAGTGTAGACAGAGGGCA----- 102  
Db 946 AlaGlnGluArgGluLysArgGluGlu-AlaPheArgAlaGlyThrLeuProAlaLeuPhe 965

QY 101 -----ACAATGGAGAAAGGAGTATCTATTATGAGCTTGGGCAATTTTACAC 57  
|||||  
Db 965 eCysserProThrMerGluLeuGlyValAspIleSerAspLeuAsnValValHis 993  
|||||  
RESULT 8  
H96815  
hypothetical protein F9K20.26 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: H96815  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Pederspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziali,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.W.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: H96815  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-325 <STO>  
A:Cross-references: GB:AE005173; NID:G3834322; PIDN:AAC83038.1; GSPDB:GN00141  
C:Genetics:  
A:Gene: F9K20.26  
A:Map position: 1

Alignment Scores:  
Pred. No.: 5.47 Length: 325  
Score: 82.50 Matches: 41  
Percent Similarity: 35.57% Conservative: 12  
Best Local Similarity: 27.52% Mismatches: 37  
Query Match: 9.82% Indels: 59  
DB: 2 Gaps: 8

US-10-057-510-1 (1-486) x H96815 (1-325)

QY 394 GGAGGGGCTTTATGCAAGTTCCTTCATGCAATGAGCCAAAGNCCTGTCG 335  
|||||  
Db 84 GlyGlySerAlaThrAlaSerProCysSerSerTyGlnProSerProCysAlaSer 103  
|||||  
QY 334 ---AACCCTTGACTGGCAATCTCTGAGAGCTCGGGACNATCCAGGGCTGCATCTC 278  
|||||  
Db 104 TyrAsnPro-----SerProGlySerSerAsn 112  
|||||  
QY 277 ATCTCAGCCCTCGTAGCCTTCT-----GCCCTC 248  
|||||  
Db 113 PheMetSerProAlaSerSerSerPheAlaAsnLeuThrSerGlyAspGlyGlnSerLeu 132  
|||||  
QY 247 CTCACAGGCTCGTCCAT-----TCCTTCAGCTCCAGCGCTCTCCACAGT 200  
|||||  
Db 133 IleProTrpLeuHisLeuSerThrThrSerSerSerAlaSerSerSerArg 152  
|||||  
QY 199 -----CCATGCTTCCTCCACCTTGC 182  
|||||  
Db 153 LeuProAsnTyrLeuTyrIleProGlyGlySerIleSerAlaProValThrProLeu 172  
|||||  
QY 181 CACATGGANACGCGTGGCCCCCAGCAGGAGGACACAGAGCTGGCAGAGAA 122  
|||||  
Db 173 SerSerProThrAlaArgThrPro-----ArgMetAsnThrAsp-----Trp 186  
|||||  
QY 121 CAGCAGTGTAGACAGGGGCAACATGGAGAGGATATCTATTATGAGCTTGGCATTT 62  
|||||  
Db 187 GlnGlnLeuAsnAsn-----SerPhe 193  
|||||  
QY 61 TCACACAGCACTTCCAGCTTTTCA 35  
|||||  
Db 194 PheValSerSerThrProProSerPro 202  
|||||

RESULT 9

T08101  
outer dynein arm docking complex protein ODA3 - Chlamydomonas reinhardtii  
C:Species: Chlamydomonas reinhardtii  
C:Date: 21-May-1999 #sequence\_revision 21-May-1999 #text\_change 08-Oct-1999  
C:Accession: T08101  
R:Koutoulis, A.; Pazour, G.J.; Wilkerson, C.G.; Inaba, K.; Sheng, H.; Takada, S.; Witmar  
J. Cell Biol. 137, 1069-1080, 1997  
A:Title: The Chlamydomonas reinhardtii ODA3 gene encodes a protein of the outer dynein  
A:Reference number: Z16354; MUID:97311076; PMID:9166407  
A:Accession: T08101  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-749 <KOU>  
A:Cross-references: EMBL:AF001309; NID:G2150040; PIDN:AAC49732.1; PID:G2150041  
C:Genetics:  
A:Gene: ODA3  
C:Function:  
A:Description: may play an important role in the precise positioning of the outer dynein

Alignment Scores:  
Pred. No.: 5.51 Length: 749  
Score: 82.50 Matches: 35  
Percent Similarity: 34.56% Conservative: 12  
Best Local Similarity: 25.74% Mismatches: 36  
Query Match: 9.82% Indels: 53  
DB: 2 Gaps: 7

US-10-057-510-1 (1-486) x T08101 (1-749)

QY 411 GCCAGGACCAACACGAGGGGGCTTCTTATGCAAGTTCCTTCATGCCATATGAGC 352  
|||||  
Db 485 AlaSerAlaGlyAlaArgGlySerAlaHisThrProAspArgAsn----- 502  
|||||  
QY 351 CAAGNCCTGTCGAAACCTTGACTGGCAATGCTCTGAGAGTCTGGGACNATCCA 292  
|||||  
Db 503 -----LysArgGlyProAlaThr 508  
|||||  
QY 291 GGGCTGCATCTTCATCCTCAGCCCTCGTAGCTGCTGCCCTCCACAGGCTCGTCC 232  
|||||  
Db 509 GlySerArgSerGlnSerProAla-----LeuValProHisSerProAlaGlyAsp 525  
|||||  
QY 231 ATTCCTTCAGCTCCA-----GGCCTCTCCACACGCTCCATGCCCTTCCACCTTGC 181  
|||||  
Db 526 LysProSerSerProLeuHisGlyThrSerProGluHisGlyHis----- 540  
|||||  
QY 180 ACATGGANACGCGTGGCCCCCAGCAGGAGGACACAGAGCTGGCAGAGAAAGAAC 121  
|||||  
Db 541 -----GluProIle-----ProGluGlyAlaGluLeuAla----- 551  
|||||  
QY 120 AGCAGTTAGACAGGGGCAACATGGAGAGGAGTATCTATTATGAGCTTGGG----- 67  
|||||  
Db 552 -----GlyGluAlaGluLeuValSerProLeuGlyAlaAsp 563  
|||||  
QY 66 -----CATTTACACGACACTTCCACGCTT 40  
|||||  
Db 564 GlyAsnThrIleGluAspGluHisPheProGluLeuProGluLeu 579  
|||||

RESULT 10

T19170  
hypothetical protein C09H6.1 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T19170  
R:Harris, B.  
submitted to the EMBL Data Library, November 1996  
A:Reference number: Z19083  
A:Accession: T19170  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1309 <WIL>  
A:Cross-references: EMBL:Z81466; PIDN:CAB03868.1; GSPDB:GN00019; CESP:C09H6.1  
A:Experimental source: clone C09H6  
C:Genetics:

A:Gene: CESP:C09H6.1  
A:Map position: 1  
A:Introns: 107/3; 434/3; 478/3; 564/2; 1020/1; 1240/3

Alignment Scores:  
Pred. No.: 5.53 Length: 1309  
Score: 82.50 Matches: 30  
Percent Similarity: 38.46% Conservative: 25  
Best Local Similarity: 20.98% Mismatches: 55  
Query Match: 9.82% Indels: 33  
DB: 2 Gaps: 4

US-10-057-510-1 (1-486) x TI9170 (1-1309)

QY 423 CCAGAACACAGAGGGCCAGGACCACAAACACGAGGAGGGCTTCT----- 382  
Db 921 ProgluLysylsArgLysGlnSerSerGluGluProAlaSerAspProGluLeuPheGly 940  
QY 381 TATCCAGTTCCCTTCATGCATATAGCAGCAAGNCTNGTGCTCGAACCTTGACTGG 322  
Db 941 SerSerGlnProThrARGlnLeuSerGluArgAlaThrArgAsnArgIleAsnTyr 960  
QY 321 GCAATGTCTCAGAGTAGTCTGGGGACNATCCAGGGGCTGCATCTTCATCTCCAGCCCTGCCGT 362  
Db 961 SerLeuLeuSerLysAsnGlySerGlyLysArgProThrProSerThrSerSerAlaAsnLeu 980  
QY 261 AGCCTTCTGCCCTCTCCAGGGCTCTCATCTCTTCACCTCAGCGGCTCTCCACAC 202  
Db 981 GluLysLeuAlaGlySerSerGlyGlyAlaSerSerGluSerProGluProAspGluSer 1000  
QY 201 GTCCATGCTCTCCAC-----CTTGCCACATGGANACAG----- 169  
Db 1001 ValGluValSerHisTrpLysIleArgThrPheLeuArgSerGluTyrGlyValLysGlu 1020  
QY 168 -----CGTGCCCCCACAGCCAGGAGGAACAAGAGCTGGCAAGAAG 124  
Db 1021 SerLeuLysCysProAspCysProTyrLysSerSerGluProAspValLeu----- 1037  
QY 123 AACAGCAGTTAGAGCAGGGGCAACAATGGAGAGGAGTAGTATCTATTATGAGTTGGGCAT 64  
Db 1038 -----GlulysHisArgTyrTyrHisMetThrLys 1047

QY 63 TTITACACCA 55  
Db 1048 ThrThrPro 1050

RESULT 11  
S09257  
homeotic protein Hox A4 - chicken  
N:Alternate names: homeotic protein Choxi-4  
C:Species: Gallus gallus (chicken)  
C>Date: 29-Jan-1993; Sequence\_revision 29-Jan-1993 #text\_change 21-Jul-2000  
C:Accession: S09257; SI0883  
R:Sasaki, H.; Yokoyama, E.; Kuroiwa, A.  
Nucleic Acids Res. 18, 1739-1747, 1990  
A>Title: Specific DNA binding of the two chicken deformed family homeodomain proteins, Q  
A:Reference number: S09256; MUID:90245562; PMID:1970866  
A:Accession: S09257  
A:Molecule type: mRNA  
A:Residues: 1-309 <SR>  
A:Cross-references: EMBL:X52670; NID:g63218; PIDN:CAA36896.1; PID:g63219  
R:Scotting, P.J.; Hewitt, M.; Keynes, R.J.  
Nucleic Acids Res. 18, 3999, 1990  
A>Title: Isolation and analysis of chick homeobox cDNA clones.  
A:Reference number: SI0883; MUID:90326535; PMID:1973835  
A:Accession: SI0883  
A:Molecule type: mRNA  
A:Residues: 207-273 <SCO>  
A:Cross-references: EMBL:X52747; NID:g63223; PIDN:CAB57949.1; PID:g6018426  
C:Function:  
A:Description: control of embryonic development by tissue- and stage-specific regulation  
C:Superfamily: homeotic protein Hox D4; homeobox homology  
C:Keywords: DNA binding; embryo; homeobox; nucleus; transcription regulation

F;210-266/Domain: homeobox homology <HOX>

```
Alignment Scores:
Pred. No.:      7.7      Length:      309
Score:          81.00    Matches:      31
Percent Similarity: 34.91% Conservative: 6
Best Local Similarity: 29.25% Mismatches: 31
Query Match:     9.64% Indels:      38
DB:              1      Gaps:       4

US-10-057-510-1 (1-486) x S09257 (1-309)
Qy 341 TGGCTGCAACCTTGACTGGGCAATGTCCTGAG-----AGTCTGGG 300
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 12 TyIleGlupProLysPheProCysGlulnTyThrGlnHisSerGlySerAlagly 31
Qy 299 ACNATCCAGGGGTGCATCTTCATCTCATGCCCTCGCTAGCCCTTCTGCCCTCTCCAGG 240
    :::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 32 SerSerAlaserTyHisProHisHis-----ProHisProHis 44
Qy 239 GCTCGTCCATCTCTCAGCTCCAGCGCTCTCCACACGTCCTCCACCTTCCACCTGCCA 180
    :::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 45 AlaProPro--ProProProProProProProHisLeuHisAlaHis----- 61
Qy 179 CATGGANACAGCCGTGCCCCCAAGCAGCAGGGAACAACGAGCTGGCAGAAAGAACA 120
    :::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 62 -----ProGlyProAlaLeuProGluTyPheProArgProArgarg----- 75
Qy 119 GCAGTTAGACGAGGGGCAACAATGGAGAAAGAGTATCTATTATGAGCTTGGGCATTTTA 60
    :::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 76 -----GluproglyTyrglnA 81

RESULT 12
AG2968
Hypothetical protein Atu3352 [imported] - Agrobacterium tumefaciens (strain C58, Dupont
C/Species: Agrobacterium tumefaciens
C/Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C/Accession: AG2968
R/Wood, D.W.; Serubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo,
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McCiel,
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A/Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A/Reference number: AB2577; MUID:21608550; PMID:11743193
A/Accession: AG2968
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-410 <KUR>
A/Cross-references: GB:AE00869; PIDN:AAU44165.1; PID:gl7741740; GSPDB:GN00187
A/Experimental source: strain C58 (Dupont)
C/Genetics:
A/Gene: Atu3352
A/Map position: linear chromosome

Alignment Scores:
Pred. No.:      7.72     Length:      410
Score:          81.00    Matches:      37
Percent Similarity: 33.76% Conservative: 16
Best Local Similarity: 23.57% Mismatches: 47
Query Match:     9.25% Indels:      57
DB:              2      Gaps:       8

US-10-057-510-1 (1-486) x AG2968 (1-410)
```

QY 125 TTTCTTGGCAGCTTCGTGTTCCCTCTGCCTTGCGGGGACGGCTGNTCCATGTGGCA 184  
||||| :|||  
Db 10 PhLeuAlaSerValMetIleProAlaGlyMetalGlyAlaThrAspLeuGluValThr 29  
||||| :|||

```

QY      185 AGTGTGAAGCCATGACGCTGTGTGGAGGAGCCCTGGAGCT----- 223
Db      30 HisTrp-----TrpThrSerGlyGluAlaAlaValAlaGluLeuAlaLysAla 47
QY      224 ---GAAGGAAATGACGAGCCCTGGGAGGAGGCG----- 253
Db      48 PheAspAlaThrGlyAsnLysTrpValAspGlyAlaLeuAlaGlySerGlyGlyThrAla 67
QY      254 -----AGAAGGCTACGC----- 265
Db      68 AspProIleMetIleSerArgIleThrGlyGlyAspProMetGlyAlaThrGlnPheAsn 87
QY      266 -----AGCGCTGAGATGACATGCACCCCTGGATNGTCCAGACTCTCAGGACATT 319
Db      88 HisGlyArgGluAlaGluGluValGlnAlaGlyLeuMetArgAspLeuSerAspVal 107
QY      320 GCCCAGTCAAGGG-----TTTCAGCCCAACNAGN---CTTGCTCATATGCG 361
Db      108 Ala-AlaLysGlyAsnTrpLysGluValIleLysProAlaSerLeuLeuAspSerCysTh 127
QY      362 CATGAAGGGGACTTGATAGAAAGCC-----CCTCCCTGGTTGTGCTCCTGGCC 412
Db      127 rIleAspGlyLysIleTyrCysAlaProValAsnIleHisSerTrpIleTrpLeuTrpLe 147
QY      413 CTCT-----GTTCTGAAAACTGG 431
Db      147 uSerAsnAlaAlaPheLysLysAlaGlyValGluValProLysAsnTrp 163

RESULT 13
D98314
hypothetical protein AGR_L 2941 [imported] - Agrobacterium tumefaciens (strain
C;Species: Agrobacterium tumefaciens
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
C;Accession: D98314
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorllo, B.;
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.;
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobac
A;Reference number: A97359; MUID:21608551; PMID:11743194
A;Accession: D98314
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-410 <KUR>
A;Cross-references: GB:AE007870; PIDN:AAK90038.1; PID:g15160015; GSPDB:GN00170
C;Genetics:
A;Gene: AGR_L 2941
A;Map position: linear chromosome

Alignment Scores:
Pred. No.: 7.72 Length: 410
Score: 81.00 Matches: 37
Percent Similarity: 33.76% Conservative: 16
Best Local Similarity: 23.57% Mismatches: 47
Query Match: 9.25% Indels: 57
DB: 2 Gaps: 8

US-10-057-510-1 (1-486) x D98314 (1-410)

QY      125 TTCTTGTCCAGCTGCTGTGTTGCTCTGCTGTGGGGCAGCGCTGTTCCATGTGGCA 184
Db      10 PheLeuAlaSerValMetIleProAlaGlyMetAlaGlyAlaThrAspLeuGluValThr 29
QY      185 AGGTGGAAGGCATGACCTGTGTGGAGGAGCGCTGGAGCT----- 223
Db      30 HisTrp-----TrpThrSerGlyGlyAlaAlaValAlaGluLeuAlaLysAla 47
QY      224 ---GAAGGAAATGACGAGCCCTGGGAGGAGGCG----- 253
Db      48 PheAspAlaThrGlyAsnLysTrpValAspGlyAlaLeuAlaGlySerGlyGlyThrAla 67
QY      254 -----AGAAGGCTACGC----- 265

```

[illegible]



Db 153 ProMetSerProThrProArgGlnPheGlyArgAlaProValProPheAlaGlnPro 172  
Qy 240 -----GGCTGCTCCATTCTTCAGCTCCAGCGCTCCT 208  
Db 173 PheGlyAlaGluProGluAlaProTTPGlySerSerProSerProProProPro 192  
Qy 207 CCACAGTC 199  
Db 193 ProProVal 195

## RESULT 15

T02345  
hypothetical protein KIAA0324 - human (fragment)  
C:Species: Homo sapiens (man)  
C>Date: 05-Mar-1999 #sequence\_revision 05-Mar-1999 #text\_change 05-Nov-1999  
C:Accession: T02345  
R:Ricke, D.O.; Bruce, D.; Mundt, M.; Doggett, N.; Munk, C.; Saunders, E.; Robinson, D.;  
re, J.; White, S.; Ueng, S.; Tatum, O.; Campbell, C.; Fawcett, J.; Deaven, L.  
submitted to the EMBL Data Library, March 1998  
A:Description: Sequencing of human chromosome 16p13.3.  
A:Reference number: Z14664  
A:Accession: T02345  
A>Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-1791 <RIC>  
A:Cross-references: EMBL:AC004493; NID:g2996648; PIDN:AAC08453.1; PID:g2996650  
C:Genetics:  
A:Map position: 16  
A:Introns: 1610/2; 1706/2  
A:Note: KIAA0324

## Alignment Scores:

Pred. No.:	7.81	Length:	1791
Score:	81.00	Matches:	52
Percent Similarity:	42.07%	Conservative:	17
Best Local Similarity:	31.71%	Mismatches:	50
Query Match:	9.64%	Indels:	45
DB:	2	Gaps:	11

US-10-057-510-1 (1-486) x T02345 (1-1791)

Qy 464 TCAGGCTTCNAAACACCGGGGNCNTANGANCAGTTTCAGAACAGGGCCAGGA 405  
Db 426 AlaGlyMetSerSerAsnGlnSerIleSerSerProValLeu--AspAlaValProArgT 445  
Qy 404 CCACAAACAGGAGGGGCTTCTTATGCAAGTTCCTCATGCCATATGAGCCAGNCC 345  
Db 445 hrProSerArgGluArgSerSerAlaSerPro-----GluMetLysAspGlyL 463  
Qy 344 TNGTGGCTCGAACCTTGACTGGGCAATGTCCTGAGAGTCTGGGACNATCCAGGGCT- 286  
Db 463 eu----ProArgThrPro-----SerArgSerArgSerGlySerProGlyLeuA 480  
Qy 285 -----GCATCTTCATCTCAGCCCTCGTAGCCTTCTGCCCTCTCCAGGGC---TCGT 234  
Db 480 rgaspGlySerGlyThrProSerArgHisSerLeuSerGlySerSerProGlyMetLysA 500  
Qy 233 CCATTCCTTCAGTCCA-----GGCTCTCTCCACAGC 201  
Db 500 spileProArgThrProSerArgGlyArgSerGluCysAspSerSerProGluPro-Lys 519  
Qy 200 TCCATGCTTCCACCTTGCCACATGGANACAGCGCTGCCCCACACAGC----- 153  
Db 520 AlaLeuProGlnThr---ProArgProArgSerProSerSerProGluLeuAsn 538  
Qy 152 -----CAGAGGGAACAGAGCTGGCAAGAA----- 126  
Db 539 AsnLysCysLeuThrProGlnArgGlu---ArgSerGlySerGluSerSerValAspGln 557  
Qy 125 -----AGACAGCAGT-----AGAGCAGGGGCAACAATGGAGAA 90  
Db 558 LysThrValAlaArgThrProLeuGlyGlnArgSerArgSerGlySerSerGlnGluLeu 577

Qy 89 GGAGTA 84  
Db 578 AspVal 579

Search completed: April 6, 2004, 17:58:32  
Job time : 24 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: April 6, 2004, 17:55:25 ; Search time 42.5 Seconds  
(without alignments)  
6006.579 Million cell updates/sec

Title: US-10-057-510-1

Perfect score: 876  
Sequence: 1 gcaagcagctatnaaaat.....aaagggaaantccccatggg 486

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1071772 seqs, 262633353 residues

Total number of hits satisfying chosen parameters: 2143544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame\_n2p\_model -DEV=xlp  
-Q=/cgn2\_1/USPTO\_spool\_P/US10057510/runat\_06042004\_141610\_14799/app\_query.fasta\_1.647  
-DB=published-Applications-AA -QFM=fastan -SUFFIX=rapb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US10057510@cgn1.1.13 @runat\_06042004\_141610\_14799  
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-XGAPOP=6 -XGAPEXT=7 -XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA:\*

1:	/cgn2_6/ptodata/1/pubaa/US07_PUBCOMB.pep.*
2:	/cgn2_6/ptodata/1/pubaa/PCT_NEW_PUB.pep.*
3:	/cgn2_6/ptodata/1/pubaa/US06_NEW_PUB.pep.*
4:	/cgn2_6/ptodata/1/pubaa/US06_PUBCOMB.pep.*
5:	/cgn2_6/ptodata/1/pubaa/US07_NEW_PUB.pep.*
6:	/cgn2_6/ptodata/1/pubaa/PCTUS_PUBCOMB.pep.*
7:	/cgn2_6/ptodata/1/pubaa/US08_NEW_PUB.pep.*
8:	/cgn2_6/ptodata/1/pubaa/US08_PUBCOMB.pep.*
9:	/cgn2_6/ptodata/1/pubaa/US09A_PUBCOMB.pep.*
10:	/cgn2_6/ptodata/1/pubaa/US09B_PUBCOMB.pep.*
11:	/cgn2_6/ptodata/1/pubaa/US09C_PUBCOMB.pep.*
12:	/cgn2_6/ptodata/1/pubaa/US09_NEW_PUB.pep.*
13:	/cgn2_6/ptodata/1/pubaa/US10A_PUBCOMB.pep.*
14:	/cgn2_6/ptodata/1/pubaa/US10B_PUBCOMB.pep.*
15:	/cgn2_6/ptodata/1/pubaa/US10C_PUBCOMB.pep.*
16:	/cgn2_6/ptodata/1/pubaa/US10_NEW_PUB.pep.*
17:	/cgn2_6/ptodata/1/pubaa/US60_NEW_PUB.pep.*
18:	/cgn2_6/ptodata/1/pubaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
-----				

1	394	45.0	74	13	US-10-057-510-2	Sequence 2, Appli
2	397	44.2	247	15	US-10-131-410-116	Sequence 116, App
3	183.5	20.9	196	15	US-10-264-237-2428	Sequence 2428, Ap
4	94.5	11.2	19695	15	US-10-084-846A-3	Sequence 3, Appli
5	87.5	10.0	1725	14	US-10-029-388-27616	Sequence 27616, A
6	87.5	10.0	19723	15	US-10-084-846A-5	Sequence 5, Appli
7	86	10.2	163	9	US-09-738-973-41	Sequence 41, Appl
8	86	10.2	163	9	US-09-854-133-41	Sequence 41, Appl
9	86	10.2	163	14	US-10-144-649A-41	Sequence 41, Appl
10	86	10.2	378	14	US-10-156-761-9745	Sequence 9745, Ap
11	84.5	10.1	196	12	US-10-425-114-52720	Sequence 52720, A
12	84.5	10.1	250	12	US-10-425-114-53402	Sequence 53402, A
13	84.5	10.1	250	12	US-10-425-114-72216	Sequence 72216, A
14	84.5	10.1	261	12	US-10-425-114-63205	Sequence 63205, A
15	84	9.6	338	14	US-10-171-311-71	Sequence 71, Appl
16	84	9.6	338	14	US-10-225-567A-110	Sequence 110, Appl
17	83.5	9.9	817	15	US-10-369-493-1813	Sequence 1813, Ap
18	83	9.9	192	15	US-10-108-260A-4083	Sequence 4083, Ap
19	83	9.9	213	14	US-10-147-170A-5	Sequence 5, Appli
20	83	9.9	217	15	US-10-108-260A-4063	Sequence 4063, Ap
21	83	9.9	221	14	US-10-147-170A-6	Sequence 6, Appli
22	83	9.9	239	9	US-09-925-299-834	Sequence 834, App
23	83	9.9	239	10	US-09-925-299-834	Sequence 834, App
24	83	9.9	239	14	US-10-106-698-5808	Sequence 5808, Ap
25	83	9.9	254	15	US-10-334-143-29	Sequence 29, Appl
26	83	9.5	542	12	US-10-276-774-1386	Sequence 1386, Ap
27	82.5	9.8	165	12	US-10-425-114-46724	Sequence 46724, A
28	82.5	9.8	213	12	US-10-220-120-415	Sequence 415, App
29	82.5	9.8	597	15	US-10-369-493-3265	Sequence 3265, Ap
30	82	9.4	119	12	US-10-424-599-180662	Sequence 180662,
31	81	9.6	371	12	US-10-425-114-71775	Sequence 71775, A
32	80.5	9.2	116	12	US-10-424-599-222962	Sequence 222962,
33	80.5	9.6	228	12	US-10-425-114-68832	Sequence 68832, A
34	80	9.5	717	14	US-10-156-761-14833	Sequence 14833, A
35	80	9.5	1265	9	US-09-862-027-80	Sequence 80, Appl
36	80	9.5	1798	16	US-10-311-034-23	Sequence 23, Appl
37	79.5	9.5	192	12	US-10-425-114-68764	Sequence 68764, A
38	79.5	9.5	193	12	US-10-425-114-67400	Sequence 67400, A
39	79.5	9.5	651	15	US-10-346-144-2	Sequence 2, Appli
40	79.5	9.5	724	15	US-10-346-144-5	Sequence 5, Appli
41	79.5	9.5	1239	12	US-10-282-122A-76629	Sequence 76629, A
42	79	9.4	157	11	US-09-864-408A-1498	Sequence 1498, Ap
43	79	9.0	5546	12	US-10-296-734-1210	Sequence 1210, Ap
44	79	9.4	19608	15	US-10-084-846A-8	Sequence 8, Appli
45	78.5	9.0	243	14	US-10-050-704-283	Sequence 283, App

#### ALIGNMENTS

RESULT 1  
US-10-057-510-2  
; Sequence 2, Application US/10057510  
; Publication No. US20020098580A1  
; GENERAL INFORMATION:  
; APPLICANT: Nandabalan, Krishnan  
; APPLICANT: Yang, Meijia  
; APPLICANT: Schulz, Vincent  
; APPLICANT: Curagen Corporation  
; TITLE OF INVENTION: MDM INTERACTING PROTEIN AND METHODS OF USE THEREOF  
; FILE REFERENCE: 15966-524 MDM US  
; CURRENT APPLICATION NUMBER: US/10/057,510  
; PRIOR APPLICATION NUMBER: USSN 09/510,252  
; PRIOR FILING DATE: 2000-02-22  
; PRIOR APPLICATION NUMBER: USSN 60/121,192  
; PRIOR FILING DATE: 1999-02-23  
; PRIOR APPLICATION NUMBER: USSN 60/122,643  
; PRIOR FILING DATE: 1999-03-03  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: Patentin Ver. 2.0

```

; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Any X can be any amino acid.
US-10-057-510-2
Alignment Scores:
Pred. No.: 4.23e-31 Length: 74
Score: 394.00 Matches: 74
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 44.98% Indels: 0
DB: 13 Gaps: 0

US-10-057-510-1 (1-486) x US-10-057-510-2 (1-74)
QY 1 GCACGACGATNAAAAATGTTTCCATGTTTATGAAAGCTGGAGTCTGGTGT 60
Db 1 AlaAargAlaTyR**LysMetPheSerMetPheMetGluLysAlaGlyLysCysTrpCys 20
QY 61 ARAATGCCAGCTCATATAGACTACTCTTCTCCATTGTTGCCCTGCTCTAACTGCT 120
Db 21 LysMetProLysLeuIleIleAspThrProPheSerIleValAlaProAlaLeuThrAla 40
QY 121 GTTCTTTTCCGACGCTTCCTGTTTCCCTCTGGCTTGTGGGCGACGGCTGTNTCCATGT 180
Db 41 ValLeuSerCysGlnLeuArgCysSerLeuTrpLeuValGlyAlaArgLeu**ProCys 60
QY 181 GCGAAGGTGAAGCATGACGTGTGGAGGAGCGCTGGAGC 222
Db 61 GlyLysValGluGlyMetAspValTrpArgArgTrpSer 74

RESULT 2
US-10-131-410-116
; Sequence 116 Application US/10131410
; Publication No. US20030235915A1
; GENERAL INFORMATION:
; APPLICANT: SPECHT, THOMAS
; APPLICANT: HINZMANN, BERND
; APPLICANT: SCHMITT, ARMIN
; APPLICANT: PILARSKY, CHRISTIAN
; APPLICANT: DAHL, EDGAR
; APPLICANT: ROSENTHAL, ANDRE
; TITLE OF INVENTION: HUMAN NUCLEIC ACID SEQUENCES FROM TISSUE OF BREAST
; FILE REFERENCE: SCH-1763
; CURRENT FILING DATE: 2002-04-25
; PRIOR FILING DATE: 2002-04-25
; PRIOR APPLICATION NUMBER: US/10131,410
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: PCT/DE99/00908
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 202
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 116
; LENGTH: 247
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-131-410-116
Alignment Scores:
Pred. No.: 2.66e-30 Length: 247
Score: 387.00 Matches: 94
Percent Similarity: 72.99% Conservative: 6
Best Local Similarity: 68.61% Mismatches: 30
Query Match: 44.18% Indels: 8
DB: 15 Gaps: 2

US-10-057-510-1 (1-486) x US-10-131-410-116 (1-247)
QY 87 TCCTTCTCCATGTTGCCCTGCTCTAACTGCTGTTCTTCTGCGACGTTGGTGTTC 146
Db 34 SerPheLeuHisAlaThrHisLeuSer---AlaGluValLysAlaLeuSerLeuPhe 52

; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Any X can be any amino acid.
US-10-057-510-1
Alignment Scores:
Pred. No.: 8.9e-10 Length: 196
Score: 183.50 Matches: 51
Percent Similarity: 69.14% Conservative: 5
Best Local Similarity: 62.96% Mismatches: 19
Query Match: 20.95% Indels: 6
DB: 15 Gaps: 1

```

```
US-10-057-510-1 (1-486) x US-10-264-237-2428 (1-196)
QY 254 AGAAGGCTACGAGGCTGAGATGAATGACCGCCCTGGATNGTCCCGAGCTCTCAG 313
Db 38 ArgArgLeuArgAlaGluaspGluaspAlaAlaProGlyTrpSer-GlnThrLeuAr 57
QY 314 GACATTGCCAGTCAAGGTTTCGAGCCACNAGNCTTGGCTCATATGGCATGAAGGGAA 373
Db 57 gThrCysProAlaGlnGlyLeuArgAlaThrGlyLeuAlaSerTrpGlyMetArgGlyse 77
QY 374 CTTCATAAAGAGCCCTCCCTGGTGTGTGGTCTCTGGCCCTCTCTTCTTGAAAACGTGNT 433
Db 77 rTriHis-ArgSerProLeuPro-AlaValValLeu-ProSerValLeuGlnThr--Al 95
QY 434 CNTAGNCCCTCGTTTTCGAAAGCCTGAGAAAGGAAAGGAAANTCCCGCATGG 486
Db 95 aLeuSerProLeuAlaLeuCysGlnAlaTrpArg-ArgAlaValProHisGly 112

RESULT 4
US-10-084-846A-3
; Sequence 3, Application US/10084846A
; Publication No. US20040006026A1
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MUHLENWEG, AGNES
; APPLICANT: TREFFZER, AXEL
; APPLICANT: BECHTHOLD, ANDREAS
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; FILE REFERENCE: 1974-005
; CURRENT APPLICATION NUMBER: US/10/084,846A
; PRIOR FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: PCT/EP01/09815
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; PRIOR FILING DATE: 2001-02-25
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: Patent in ver. 3.2
; SEQ ID NO 3
; TYPE: PPT
; ORGANISM: Streptomyces viridochromogenes
; FEATURE:
; OTHER INFORMATION: Protein 1: amino acid sequence encoded by coding strand 1.
; OTHER INFORMATION: Start codon: gga, Start position: nucleotide 1.
US-10-084-846A-3

Alignment Scores:
Pred. No.: 1.92 Length: 19695
Score: 94.50 Matches: 40
Percent Similarity: 42.52% Conservative: 14
Best Local Similarity: 31.50% Mismatches: 37
Query Match: 11.25% Indels: 36
DB: 15 Gaps: 6

US-10-057-510-1 (1-486) x US-10-084-846A-3 (1-19695)
QY 423 CCAGAACAGAGGCGCAGGACCCACACAGGAGGGGCTTCTATGCAAGTCCCTTCA 364
Db 766 ProThrSerProAlaSerThrThrArg-----ProSerAenSerSerProSer 783
QY 363 TGCCATATGAGCCCAAGNCTGTGGCTCGAACCCCTTGACTGGGCAATGTCTCTGAGAGTCT 304
Db 784 -----AlaArgThr-----SerProAlaSer 790
QY 303 GGGGACNATCCAGGGGTGATCTTCATCTCCACCCCTGGTAGCTTCCCTCTCTCC 244
Db 791 ProProSerProAlaSerAlaSerThrProSerAlaCysArgProThrArgProAlaArg 810
QY 243 CAGGGCTCGTCCANT---CCTTCAGCTCCAGCGCTCTCCCA----- 205
Db 811 ArgasnProAlaProGlyProSerArgAlaProProProProProProProProProAlaSer 830

US-10-029-386-27616
; Sequence 27616, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 27616
; LENGTH: 175
; TYPE: PPT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC013648.3
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.6
; OTHER INFORMATION: SWISSPROT HIT: Q92671, EVALUE 2.40e+00
US-10-029-386-27616

Alignment Scores:
Pred. No.: 4.28 Length: 175
Score: 87.50 Matches: 18
Percent Similarity: 52.63% Conservative: 2
Best Local Similarity: 47.37% Mismatches: 17
Query Match: 9.99% Indels: 1
DB: 14 Gaps: 1

US-10-057-510-1 (1-486) x US-10-029-386-27616 (1-175)
QY 146 CCCTCTGGCTTGTGGGGCAC---GGCTGTNTCCATGTGGCAAGTGGAGGCATGGAGC 202
Db 106 ProSerGlyLeuTrpGlyAlaLeuGlySerProGlyArgAlaThrTrpArgGluTrpGlu 125
QY 203 TGTGAGAGGGCGCTGGAGCTGAAGGAATGGACGACCTGGGAGGAGGCGAGA 256
Db 126 GlyAlaGlyGlyGlnGlyValGlyLeuAlaAlaIleTrpGluAlaGluArg 143

RESULT 6
US-10-084-846A-5
; Sequence 5, Application US/10084846A
; Publication No. US20040006026A1
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MUHLENWEG, AGNES
; APPLICANT: TREFFZER, AXEL
; APPLICANT: BECHTHOLD, ANDREAS
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; FILE REFERENCE: 1974-005
```

```
; CURRENT APPLICATION NUMBER: US/10/084,846A
; CURRENT FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: PCT/EP01/09815
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; PRIOR FILING DATE: 2001-02-25
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: Patentin Ver. 3.2
; SEQ ID NO 5
; TYPE: PRT
; ORGANISM: Streptomyces viridochromogenes
; FEATURE:
; OTHER INFORMATION: Protein 3: amino acid sequence encoded by coding strand 1.
; OTHER INFORMATION: Start codon: atc, Start position: nucleotide 3.
US-10-084-846A-5

Alignment Scores:
Pred. No.: 9 8 Length: 19723
Score: 87.50
Percent Similarity: 33.58%
Best Local Similarity: 30.60%
Query Match: 9.99%
Indels: 43
DB: 15 Gaps: 10

US-10-057-510-1 (1-486) x US-10-084-846A-5 (1-19723)
QY 102 TGCCCTGCTCAACTGCTGT-----TCTTCTCCAGCTTCGTTGTTCCCTCTGGCT 155
Db 13931 CysAlaGlySerValCysCysA:GThrGlyArgArgProGlyAlaSerPro----- 13948
QY 156 TGTGGGGGCGCGCTGNTCCATGTGCGAAGTGGGAAGGCATGGAGCTGTGGAGGCGG 215
Db 13949 CysGlyCysSerAlaIleA:GMetTpsSerGlyThrMetSer----- 13962
QY 216 CTGAGCTGAAGGAATGGAAGAGCCCTGGG-----AGAGGGCAGAAGGCTACGCA 266
Db 13963 -----ThrSerThrThrSerProGlySerCysSerArgCysAlaAlaGlySerAla 13979
QY 267 GGGCTGAGGTAGAGTACAGCCCTGGATNGTCCAGACTCTCAGGACATTGCCCACT 326
Db 13980 AlaAla-Gly-----CysProProThrThrSerAlaA:GAla-----CysProLe 13994
QY 327 CAAGGGTTCGAGCCACNAGNCTTGGCTCATATGGCATGAAGGGGAACATTGCATAGAAG 386
Db 13994 u-----ArgAlaAlaCysTrpArgSe 14001
QY 387 C---CCCTCCCT-----GTTGTGGTCTGGCCCTCTGTTCTG 422
Db 14001 rCysProValProAlaAlaProSerArgCysAlaGlySerGlyAspGlyPro----- 14018
QY 423 GAAAACTGGNTCTAGNCCCTCGTGGTTTNGCAAAGCCT 462
Db 14019 -SerThrGlyArgAlaThrGlyTrpMetSerHisGlyPro 14031

RESULT 7
US-09-738-973-41
; Sequence 41, Application US/09738973
; Patent No. US20020110563A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Fling, Steven P.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Algate, Paul A.
; APPLICANT: Secrist, Heather
; APPLICANT: Indrias, Carol Yoseph
; APPLICANT: Benson, Darin R.
; APPLICANT: Elliot, Mark
; APPLICANT: Mannion, Jane
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
```

```
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.475C9
; CURRENT APPLICATION NUMBER: US/09/738,973
; CURRENT FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 587
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 41
; LENGTH: 163
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-738-973-41

Alignment Scores:
Pred. No.: 5 99 Length: 163
Score: 86.00
Percent Similarity: 43.22%
Best Local Similarity: 33.05%
Query Match: 9
DB: 5 Gaps: 5

US-10-057-510-1 (1-486) x US-09-738-973-41 (1-163)
QY 462 AGGCTTTGCHAAAAACCCAGGGGNCNTANGANCAGTTTTCAGAACAGAGGGCCAGGACC 403
Db 11 ArgLeuSerLeuArgPro-----SerIleProGluLeuSerGluA:GThr 25
QY 402 ACAACAGGAGGGGCTTCTTATGCAAGTTCCTTCATGCCCATATGACCCAGNCCCTN 343
Db 26 SerArgProCysA:GAlaSerProAlaSerLeuProSerGlnHisThrSerSerProAla 45
QY 342 GTGGCTGAACCTTGACTGGGCAATGCC-----TGAGAGTCTGGGACGNATCCA 292
Db 46 GlnAlaArgValArgAsnLeuAlaGlnSerThrPheProLeuAlaIleGlnGluThrPro 65
QY 291 GGGGCTGCATCTTCTCATCTCAGCCCTCGTAGCCTTCTGCCCTCTCCAGGGGCTCGTCC 232
Db 56 GlyArgAlaProAlaHisAlaProLeuSerSerPheValProGlyValGlyGlyArgSer 85
QY 231 ATTCTCTCAGCTCCAGGCGCTCTCCACAGCTCCATCCCTTCCACCTTGCACATGGA-- 174
Db 86 -----ProAlaSer-ValGlyIleSerAlaProGlyGlyGlyProSerGlyAl 101
QY 173 -NACAGCGTCCGCC-----ACAAGCAGAGGGAACACAGAAAGCTGG 132
Db 101 aAlaAlaIleProLeuGluLeuThrGlnSerArgValGlnIleTrp 118

RESULT 8
US-09-854-133-41
; Sequence 41, Application US/09854133
; Publication No. US20020183499A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Henderson, Robert A.
; APPLICANT: Benson, Darin R.
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.475C10
; CURRENT APPLICATION NUMBER: US/09/854,133
; CURRENT FILING DATE: 2001-05-11
; NUMBER OF SEQ ID NOS: 735
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 41
; LENGTH: 163
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-854-133-41

Alignment Scores:
Pred. No.: 5 99 Length: 163
Score: 86.00
Percent Similarity: 43.22%
```

```

Best Local Similarity: 33.05%      Mismatches: 50
Query Match:      10.24%      Indels: 17
DB:                9          Gaps: 5

US-10-057-510-1 (1-486) x US-09-854-133-41 (1-163)

QY 462 AGGCTTTCGNAACACAGGGGCTGAGCCAGGAGCC 403
DB 11 ArgLeuSerLeuArgPro-----SerIleProGluLeuSerGluArgThr 25

QY 402 ACAACAGGAGGGGGCTTCTTATGCAAGTTCCTTCATGCGCATATGAGCCAAAGNCCN 343
DB 26 SerArgProCysArgAlaSerProAlaSerLeuProSerGlnHisThrSerSerProAla 45

QY 342 GTGGCTCGAACCTTGACTGGCCATGTC-----TGAGCTCTGGGACNATCCA 292
DB 46 GlnAlaArgValArgAsnLeuAlaGlnSerThrPheProLeuAlaGlnGluThrPro 65

QY 291 GGGGCTGCATCTTCCTCAGCCCTGCTAGCCTTCTGCTCCCTCCACAGGCTCGTCC 232
DB 66 GlyArgAlaProAlaHisAlaProLeuSerPheValProGlyValGlyArgSer 85

QY 231 ATTCCTTACGTCAGCGGCTCTCACACGCTCCATGCTTCCACCTTGCCACATGGA-- 174
DB 86 -----ProAlaSer-ValGlyIleSerAlaProGlyGlyProSerGlyAl 101

QY 173 -NACAGCGTGCCTCC-----ACAAGCCAGAGGGAACAACGAGCTGG 132
DB 101 aAlaAlaLysIleProLeuGluLeuThrGlnSerArgValGlnLysIleTrp 118

RESULT 10
US-10-156-761-9745
; Sequence 9745, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 9745
; LENGTH: 378
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-9745

Alignment Scores:
Pred. No.: 6.94      Length: 378
Score: 86.00      Matches: 37
Percent Similarity: 35.33%      Conservative: 16
Best Local Similarity: 24.87%      Mismatches: 37
Query Match: 10.24%      Indels: 60
DB: 14          Gaps: 7

US-10-057-510-1 (1-486) x US-10-156-761-9745 (1-378)

QY 342 GTGGCTCGAACCTTGACTGGCAATGCTCTGAGAGTCTGGGACNATCCAGGGGCTGCA 283
DB 5 ValAlaArgGlnLeuGluTrp-----GlyThrVal 14

QY 282 TCTTCATCTCAGCCCTGCTAGCCTTCTGCTCCCTCCAGGCTCGTCCATT----- 229
DB 15 SerSerSerThrAlaAlaSerGlyIleAspProIleAlaGluAlaProLeuSerLeu 34

QY 228 -----CCTTCAGCTCCAGCG-----CCTCCTCCA 205
DB 35 CysAlaArgAlaProHisValProAlaAspArgLeuValAlaGluMetValProPro 54

QY 204 CACGTCCATGCTTCCACCTTGCACATGAGNACAGCGCTGCCCCCAAGCCAGAGGGA 145
DB 55 ArgPheAspSerValArgPheAlaThrTyrlle---ProAspProAsnGlnProSerGln 73

QY 144 ACA-----ACGAGCTGGCAAGAAAGACAGCAGT 115
DB 74 ThrGluAlaValArgValLeuGluAspPheAlaThrGlyLeuGlyAlaHisAlaSer 93

QY 114 TAGAGC-----AGGGGCAACAATGGAGAA 91
DB 114 TAGAGC-----AGGGGCAACAATGGAGAA 91

```

Db 94 GlySerGlyArgArgLeuPheGlyPheGlyLysAlaProLysThrProAlaGlyPro 113  
QY 90 AGAGATATCTATTATGAG-----CTT 70  
Db 114 ArgGlyValTyrLeuAspGlyGlyTyrGlyValGlyLysThrHisLeuLeuAlaSerLeu 133  
QY 69 GGCATTTTACACAGCACTTCCAGCCCT 40  
Db 134 TrpHisAlaThrProAlaGluProGluLeu 143

RESULT 11  
US-10-425-114-52720  
; Sequence 52720, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E  
; APPLICANT: Tabaska, Jack E  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(5313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 52720  
; LENGTH: 196  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: 700453413\_FLI.pep  
US-10-425-114-52720

Alignment Scores:  
Pred. No.: 8.77 Length: 196  
Score: 84.50 Matches: 47  
Percent Similarity: 37.57% Conservative: 21  
Best Local Similarity: 25.97% Mismatches: 63  
Query Match: 12 Indels: 50  
Gaps: 7  
DB:

US-10-057-510-1 (1-486) x US-10-425-114-52720 (1-196)

QY 480 GGGANTTCTCTTCTTCAGCTTTCGNAACCCAGGGGGNCTANGANCAGTTTCCA 421  
Db 12 GlyTrpTrpTyrPheThrArgAlaCys-----Pro 21  
QY 420 GAACAGAGGGCCAGGACACCAACAGGAGGGGCTTCTTATGCAAGTTCCTTCATGC 361  
Db 22 GluProSerThrSerThrAlaAlaSerProSerAlaAlaPro----- 39  
QY 360 CATATGAGCCAAAGNCTN-----GTGCTCGAACCTTACTGGGCAATGCTCTGAGAG 307  
Db 40 ---LeuSerArgSerLeuTyrAlaValProSerThrGlyAlaProSerAlaAsp 58  
QY 306 TCTGGGACNATCAGAGGGCTGCATCTTCATCTCCCTCAGCCCTG----- 265  
Db 59 ArgGlySerArgSerGlyThrGlyGlySerValThrAlaLeuAlaAspAlaAla 78  
QY 264 -----CGTACCTTCTGCCCTCTCTCC 244  
Db 79 AlaPhePheLeuArgAlaProArgPheLeuGlyProThrArgAspLeuIleArgSerThr 98  
QY 243 CAGGCTCGTCCATTCCTTTCAGCTCCAGCGCTCTCTC----- 207  
Db 99 ArgSerSerLysSerSerAlaValAla-ValLeuValAlaAlaGlyAlaGlySerTr 118  
QY 206 -----CACAGTCCATGCTTCCACCTTGCACATGGANACAGCCCTGCCCC 160  
Db 118 pSerAlaAlaAlaValSerSerProGluAlaAlaArgGlyGlyAsnGlySer 138

QY 159 CACAAGCCAGAGGGAACACAGAGCTGGCAAGAAAGAACAGCAGTATAGACAGGGGCAAC 100  
Db 138 rSerSerArgAlaArgCysArgGlyGluA-ArgAlaValGlyAlaLysGly----- 156  
QY 99 AATCGAAGAAAGGAGTATCTATTATGAGCTTGGGC---ATTTTACACAGCACTTCCAGC 43  
Db 157 -LeuGluSerGlyValAlaIleAlaProLeuGlyGluCysThrHisGlnHisPheMetal 176  
QY 42 C 42  
Db 176 a 176

RESULT 12  
US-10-425-114-53402  
; Sequence 53402, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E  
; APPLICANT: Tabaska, Jack E  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(5313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 53402  
; LENGTH: 250  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: 700235435\_FLI.pep  
US-10-425-114-53402

Alignment Scores:  
Pred. No.: 9.15 Length: 250  
Score: 84.50 Matches: 38  
Percent Similarity: 41.46% Conservative: 13  
Best Local Similarity: 30.89% Mismatches: 58  
Query Match: 10.06% Indels: 14  
Gaps: 5  
DB:

US-10-057-510-1 (1-486) x US-10-425-114-53402 (1-250)

QY 425 TTCCAGAACAGAGGGCCAGGACCAACAGGAGGGGCTTCTTATGCAAGTTCCTT 366  
Db 108 PheProA-Arg-ArgProGlnAlaProGlnProGlyCysAlaLeu-----ProVal 123  
QY 365 CATGCCATATGAGCCAAAGNCTNG-----TGCTCGAACCTTGTACTGGG 321  
Db 124 HisGluIleArgAlaArgAlaArgGlnAlaArgGlnProHisLeuHisProLeuHisGln 143  
QY 320 CAATCT-----CCTGAGAGTCTGGGACNATCCAGGGCTGCATCTTCATCCTCAGGCC 267  
Db 144 ProCysGlyAlaGlyAlaAlaLeuProHisHisAlaArgLeuHisCysHisLeuArgPro 163  
QY 266 TCGTAGCCCTTCTGCCCTCTCTCCAGGGCTCGTCATCTCCTTACCTCAGCGCTCTCTC 207  
Db 164 ArgGlnGlnGlnProLeuProHisAlaArg---ValArgArgLeuProArgGlnVal 182  
QY 206 CACAGCTCCATGCCCTTCCACCTTGCACATGGANACAGCCGTGCCCCCAAGCCAGAGG 147  
Db 183 HisLeuA-ArgAlaArgGlyArgAlaProHis---HisHisArgAlaProLeuArgHisGln 201  
QY 146 GAACACAGAAAGTGGCAAGAAAGAACAGCAGTATAGCAGGGGCAACATATGAGAAAGGA 87  
Db 202 GlnGluArg-LeuGlyValArgArgHisGlnGlnValSerGlyA-ArgGlyArgGlnGln 221  
QY 86 GTATCTA 80  
:::|

Db 221 yHisLeu 223

## RESULT 13

US-10-425-114-72216

; Sequence 72216, Application US/10425114

; Publication No. US20040034888A1

; GENERAL INFORMATION:

; APPLICANT: Liu, Jingdong

; APPLICANT: Zhou, Yihua

; APPLICANT: Kovalic, David K.

; APPLICANT: Screen, Steven E

; APPLICANT: Tabaska, Jack E

; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

; FILE REFERENCE: 38-21(53313)B

; CURRENT APPLICATION NUMBER: US/10/425,114

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 73128

; SEQ ID NO 72216

; LENGTH: 250

; TYPE: PRT

; ORGANISM: Zea mays

; FEATURE:

; OTHER INFORMATION: Clone ID: LIB3245-492-F2\_FLI.pep

US-10-425-114-72216

## Alignment Scores:

Pred. No.: 9,15 Length: 250  
Score: 84.50 Matches: 38  
Percent Similarity: 41.46% Conservative: 13  
Best Local Similarity: 30.88% Mismatches: 58  
Query Match: 10.08% Indels: 14  
DB: 12 Gaps: 5

US-10-057-510-1 (1-486) x US-10-425-114-72216 (1-250)

QY 425 TTCAGACAGAGGCGCCAGGACCAACACGAGGGGGCTTCTATGCAAGTTCCCTT 366

Db 108 PheProArgGProGlnAlaProGlnProGlyCysAlaLeu-----ProVal 123

QY 365 CATGCCATATGAGCAAGNCTNG-----TGCTCGAACCTTCACTGGG 321

Db 124 HisGluIleArgAlaArgAlaArgGlnAlaArgGlnProHisLeuHisProLeuHisGln 143

QY 320 CAATCT-----CCTGAGCTGGGACNATCCAGGGGCTGCTCTTCATCTCAGCCC 267

Db 144 ProCysGlyAlaGlyAlaAlaLeuProHisAlaArgLeuHisCysHisLeuArgPro 163

QY 266 TGCGTAGCCTTCGCTCCCTCCAGGGCTGCTCCATTCCTTCAGCTCCAGCGCTCCTC 207

Db 164 ArgGlnGlnGlnProLeuProHisAlaArg---ValArgGLeuProArgGlnVal 182

QY 206 CACAGCTGCTCCCTTCACCTGCGCATGAGNACAGCCGTCGCCCAACAGCCAGG 147

Db 183 HisLeuArgAlaArgGlyArgAlaProHis---HisHisArgAlaProLeuArgHisGln 201

QY 146 GAACAACAAGTGGCAAGAAGACACAGCTAGAGCGGGGCAACAATGGAGAAAGG 87

Db 202 GlnGluArg-LeuGlyValArgArgHisGlnGluValSerGlyArgGlyArgGlnGln 221

QY 86 GTATCTA 80

Db 221 yHisLeu 223

## RESULT 14

US-10-425-114-63205

; Sequence 63205, Application US/10425114

; Publication No. US20040034888A1

; GENERAL INFORMATION:

; APPLICANT: Liu, Jingdong

; APPLICANT: Zhou, Yihua

; APPLICANT: Kovalic, David K.

; APPLICANT: Screen, Steven E  
; APPLICANT: Tabaska, Jack E  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 63205  
; LENGTH: 261  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: UC-ZMFLB73008E02\_FLI.pep  
US-10-425-114-63205

## Alignment Scores:

Pred. No.: 9,22 Length: 261  
Score: 84.50 Matches: 30  
Percent Similarity: 38.60% Conservative: 14  
Best Local Similarity: 26.32% Mismatches: 33  
Query Match: 10.06% Indels: 37  
DB: 12 Gaps: 6

US-10-057-510-1 (1-486) x US-10-425-114-63205 (1-261)

QY 417 CAGAGGCCAGGACCAACACAGGGGGGCTTCTTAT-----GCAAGTTCC 370

Db 43 ArgArgArgThrAlaProArgGlyProSerTrpArgGlyProAlaSerGlyAla 62

QY 369 CTTTCATGCCATATGAGCAAGNCTNGTGGTCGACCCCTTGACTGGGCAATGCTCGA 310

Db 63 ProArgCysSer-----ProTrpSerArgAla 72

QY 309 GAGTCTGGGACNATCCAGGGGCTGCATCTTCATCTCAGCCCTGGTAGCCTTCTGCC 250

Db 73 ValSerGly-----Thr 76

QY 249 TCTCCAGGCTCGTCCATTCTTCAGCTCCAGGCGCTCTCCACACGCTCCATGCTTC 190

Db 77 ThrAlaSerGlySerArgAlaProAlaGlyProAlaProTrpPro-----SerSer 93

QY 189 CACTTTCACATGGANACAGCGCTGCCCCCAAGCCAGAGGGAAACACAGAGTGGCA 130

Db 94 ProGluAlaAlaTrpArgArgProSerProProArgProArgGly---HisArgValAla 112

QY 129 AGAAGAACAGCATGTAGAGCAGGGGCAACAATGGAGAAAG 88

Db 113 ArgAlaGly-----ThrArgGlyArgGlyAlaArg 123

## RESULT 15

US-10-171-311-71

; Sequence 71, Application US/10171311

; Publication No. US20030087270A1

; GENERAL INFORMATION:

; APPLICANT: Schlegel, Robert

; APPLICANT: Chen, Yan

; APPLICANT: Zhao, Xumei

; APPLICANT: Monahan, John

; APPLICANT: Kamatkar, Shubhangi

; APPLICANT: Glatt, Karen

; APPLICANT: Gannavarapu, Manjula

; APPLICANT: Hoerish, Sebastian

; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR

; FILE REFERENCE: MRI-035

; CURRENT APPLICATION NUMBER: US/10/171,311

; CURRENT FILING DATE: 2002-06-12

; PRIOR APPLICATION NUMBER: US 60/298,159

; PRIOR FILING DATE: 2001-06-13

; PRIOR APPLICATION NUMBER: US 60/298,155



;  
; PRIOR FILING DATE: 2001-06-13  
; PRIOR APPLICATION NUMBER: US 60/335,936  
; PRIOR FILING DATE: 2001-11-14  
; NUMBER OF SEQ ID NOS: 238  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 71  
; LENGTH: 338  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-171-311-71

Alignment Scores:  
Pred. No.: 10.8 Length: 338  
Score: 84.00 Matches: 54  
Percent Similarity: 37.23% Conservative: 16  
Best Local Similarity: 28.72% Mismatches: 56  
Query Match: 9.59% Indels: 62  
DB: 14 Gaps: 11

US-10-057-510-1 (1-486) x US-10-171-311-71 (1-338)

```
QY 42 GGCTGGAGTGTGTGTGTAATGCCAAGCTCATATAGATACCTCT---TTCTCCAT 98
Db 97 GlyTTPProValLeu-----AlaGln-LeuAlaValGlySerAlaLeuPheSerI 113
QY 99 TGTGTCCTCTGCTTAACGTGCTGCTTCT-----TGCACGCT 137
Db 113 eValValProValLeuAlaProGlyLeuGlySerThrArgSerSerAlaLeuCysSerLe 133
QY 138 TCGTTGTTCCCTCTGCTT----- 156
Db 133 uGlyTyrCysValTyrGlySerAlaPheAlaGlnAlaLeuLeuLeuGlyCysHisAl 153
QY 157 ----GTGGGGGACGGCTGTTCCATGTGGCAAGTGGAGGCGATGGACGTGTGGAGGAG 212
Db 153 aSerLeuGlyHisArgLeuGly---AlaGlyGlnValProGlyLeu-ThrLeuGlyLeu 172
QY 213 GCCTCGAGCTGAAGCA-----ATGCACGAGCCCTGGGAGGAGGCGCAGAGGC 260
Db 172 hrValGlyIleTrpGlyValAlaAlaLeuLeu-ThrLeuProValThrLeuAlaSerGly 191
QY 261 TAGCGAGG-----CTGAGGATGAGATGACGCCCTCGATNGTCCC 302
Db 192 AlaSerGlyGlyLeuCysThrLeuIleTyrSerThrGluLeuLeuAlaLeuGlnAlaThr 211
QY 303 CAGACTCTCAGACATTCGCCAGTCAAGGTTTCGAGCCACNAGNCTT---GGCTCATAT 359
Db 212 HisThrValAlaCysLeuAlaIlePheValLeuLeuProLeuGlyLeuPheGlyAlaLys 231
QY 360 GGCATGAAGGGGAACCTTGTCATAGAAGCCCTCCCTGG-----TTGTGTCCTGG 410
Db 232 GlyLeuLysLysAlaLeuGlyMetGlyProGlyProTyrMetAsnIleLeuTrpAlaTrp 251
QY 411 CCCTCTGTCTGGAACCTGGNTCAGTCCCTGTTTNGCAAGCCTGAGAAAG 470
Db 252 -PheIlePheTrp-----Trp----- 256
QY 471 GGAANTCCCATGGG 486
Db 257 -----ProHisGly 259
```

Search completed: April 6, 2004, 18:01:08  
Job time : 56.5 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: April 6, 2004, 17:53:20 ; Search time 19.5 Seconds  
(without alignments)

2573.356 Million cell updates/sec

Title: US-10-057-510-1

Perfect score: 876

Sequence: 1 gcacgagcgtatnaaaaaat.....aaaggggaantccccatggg 486

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 778828

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp  
-O=/cgn2\_1/USPTO\_spool\_p/US10057510/runat\_06042004\_141609\_14770/app\_query.fasta\_1.647  
-DB=Issued Patents AA -QWMT=faetan -SUFFIX=rai -MINMATCH=0\_1 -LOOPC1=0  
-LOOEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pco -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10057510\_QCGN\_1\_1\_27 @runat\_06042004\_141609\_14770 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	394	45.0	74	4	US-09-510-252-2
2	95	11.3	195	4	US-09-252-991A-20532
3	87.5	10.4	581	4	US-09-252-991A-20966
4	86	10.2	163	4	US-09-370-838-41
5	86	9.8	446	4	US-09-252-991A-31114
6	82.5	9.8	199	4	US-09-252-991A-25623
7	82.5	9.8	320	4	US-09-252-991A-24634
8	82.5	9.8	447	4	US-09-252-991A-32755
9	82.5	9.4	1652	4	US-09-627-650B-1
10	82.5	9.4	1652	4	US-09-436-063C-1
11	82.5	9.4	1917	4	US-09-627-650B-5
12	82.5	9.4	1917	4	US-09-436-063C-5

13	82.5	9.4	2508	4	US-09-627-650B-7	Sequence 7, Appli
14	82.5	9.4	2508	4	US-09-436-063C-7	Sequence 7, Appli
15	82.5	9.4	2544	4	US-09-627-650B-3	Sequence 3, Appli
16	82.5	9.4	2544	4	US-09-436-063C-3	Sequence 3, Appli
17	82.5	9.4	2601	4	US-09-627-650B-9	Sequence 9, Appli
18	82.5	9.4	2601	4	US-09-436-063C-9	Sequence 9, Appli
19	82	9.8	127	4	US-09-252-991A-36902	Sequence 26902, A
20	82	9.8	150	4	US-09-252-991A-31728	Sequence 31728, A
21	80.5	9.6	206	4	US-09-252-991A-34085	Sequence 24085, A
22	80	9.5	204	4	US-09-252-991A-46183	Sequence 26183, A
23	79.5	9.1	299	3	US-09-188-930-192	Sequence 192, App
24	79.5	9.5	339	4	US-09-252-991A-18253	Sequence 18253, A
25	79	9.0	147	4	US-09-252-991A-32565	Sequence 32565, A
26	79	9.0	176	4	US-09-252-991A-35290	Sequence 25290, A
27	79	9.4	229	4	US-09-252-991A-36254	Sequence 26254, A
28	78.5	9.3	202	4	US-09-252-991A-32054	Sequence 32054, A
29	78.5	9.0	299	3	US-09-188-930-332	Sequence 332, App
30	78.5	9.0	299	4	US-09-312-283C-192	Sequence 192, App
31	78.5	9.0	299	4	US-09-312-283C-332	Sequence 332, App
32	78.5	9.3	347	4	US-09-252-991A-19498	Sequence 19498, A
33	77.5	9.2	144	4	US-09-252-991A-31237	Sequence 31237, A
34	77.5	8.8	340	4	US-09-252-991A-19460	Sequence 19460, A
35	77.5	9.2	487	4	US-09-252-991A-29392	Sequence 29392, A
36	77.5	8.8	519	4	US-09-252-991A-19734	Sequence 19734, A
37	77.5	9.2	527	4	US-09-370-838-216	Sequence 216, App
38	77.5	9.2	777	4	US-09-252-991A-20378	Sequence 20378, A
39	77	8.8	168	4	US-09-252-991A-32502	Sequence 32502, A
40	77	9.2	422	4	US-09-252-991A-30314	Sequence 30314, A
41	77	9.2	447	4	US-09-252-991A-25916	Sequence 25916, A
42	76.5	8.7	360	4	US-09-252-991A-32452	Sequence 32452, A
43	76.5	9.1	1729	4	US-09-553-690-2	Sequence 2, Appli
44	76	9.0	136	4	US-09-252-991A-24892	Sequence 24892, A
45	76	9.0	148	4	US-09-252-991A-28538	Sequence 28538, A

#### ALIGNMENTS

RESULT 1  
US-09-510-252-2  
; Sequence 2, Application US/09510252  
; Patent No. 6372490  
; GENERAL INFORMATION:  
; APPLICANT: Nandabalan, Krishnan  
; APPLICANT: Yang, Meijia  
; APPLICANT: Schulz, Vincent  
; APPLICANT: Curagen Corporation  
; TITLE OF INVENTION: MDM INTERACTING PROTEIN AND METHODS OF USE THEREOF  
; FILE REFERENCE: 15966-524 MDM US  
; CURRENT APPLICATION NUMBER: US/09/510,252  
; CURRENT FILING DATE: 2000-02-22  
; PRIOR APPLICATION NUMBER: USSN 60/121,192  
; PRIOR FILING DATE: 1999-02-23  
; PRIOR APPLICATION NUMBER: USSN 60/122,643  
; PRIOR FILING DATE: 1999-03-03  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 74  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: Any X can be any amino acid.  
US-09-510-252-2

Alignment Scores: 2,26e-36 Length: 74  
Pred. No.: 394.00 Matches: 74  
Score: 100.00% Conservative: 0  
Percent Similarity: 100.00% Mismatches: 0  
Best Local Similarity: 100.00% Indels: 0  
Query Match: 44.98% Gaps: 0  
DB: 4

US-10-057-510-1 (1-486) x US-09-510-252-2 (1-74)

QY 1 GCACGAGCCTATNAAGAAATGTTTCCATGTTTATGGAAGGCTGGAGTGTGGTGT 60  
Db 1 AlaaGalatyr\*\*LysMetPheSerMetPheMetGluLysAlaGlyLysCysTrpCys 20  
QY 61 AAAATGCCCAAGCTCATATAGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 120  
Db 21 LysMetProLysLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 40  
QY 121 GTCTTTTGTCCAGCT 180  
Db 41 ValLeuSerCysGlnLeuArgCysSerLeuTrpLeuValGlyAlaArgLeu\*\*ProCys 60  
QY 181 GCGAGGTGGAGGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 222  
Db 61 GlyLysValGluGlyMetAspValTrpArgArgArgArgArgArgArgArgArgArg 74

## RESULT 2

US-09-252-991A-20532  
; Sequence 20532, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 20532  
; LENGTH: 195  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-20532

Alignment Scores:  
Pred. No.: 0.0229 Length: 195  
Score: 95.00 Matches: 41  
Percent Similarity: 39.86% Conservative: 18  
Best Local Similarity: 27.70% Mismatches: 61  
Query Match: 11.31% Indels: 28  
DB: Gaps: 6

US-10-057-510-1 (1-486) x US-09-252-991A-20532 (1-195)

QY 450 AAACGAGGGGCTGANGCAGTTTCCAGACAGAGGGCCAGGACACACAGG--- 394  
Db 50 LysProAlaProSerAlaProThrProSerTrpSerAlaArgProThrArgCys 69  
QY 393 -----GAGGGGCTTCT-----TATGCAAGTTCCCTTCATGCCAT 358  
Db 70 ArgProArgGlyArgSerProSerThrCysTrpArgTrpIleAlaProProAlaCysArg 89  
QY 357 ATGAGCCAGNCTGTGCTGAAACCTTGACTGGCAATGCTCTGAGTCTGG--- 301  
Db 90 SerCysSerThrThrIleArgGlyTrpAlaTrpArgTrpTrpArgSerSerPro 109  
QY 300 GACNATCCAGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 250  
Db 110 ProSerProAlaAlaAlaThrSerSerArgSerArgLysAlaProAlaProArgGln 129  
QY 249 -----TCCTCCAGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 211  
Db 130 CysThrAlaTrpProAlaAlaThrArgGlySerProCysProAlaAlaGlyThrThrArg 149  
QY 210 CTTCCACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 151  
Db 150 Pro-TrpSerSerSerProGlyAlaArgAlaValGlySerAlaProAlaProThrSerSe 169

QY 150 GAGGGACACAGGAGC-----TGCAGAAAGAACACAGCATTTAGG 109  
Db 169 rLeuAlaAenThrSerProCysThrArgProAlaTrpSerArgThrThrSerProArgAl 189  
QY 108 AGGGGCAACAATCGAGAAAG 89  
Db 189 aGlyAlaSer-TrpArgArg 195

## RESULT 3

US-09-252-991A-20966  
; Sequence 20966, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 20966  
; LENGTH: 581  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-20966

Alignment Scores:  
Pred. No.: 0.233 Length: 581  
Score: 87.50 Matches: 39  
Percent Similarity: 34.72% Conservative: 11  
Best Local Similarity: 27.08% Mismatches: 51  
Query Match: 10.42% Indels: 43  
DB: Gaps: 6

US-10-057-510-1 (1-486) x US-09-252-991A-20966 (1-581)

QY 471 CTTTCTTCAGGCTTTTCNAAACACAGGGGGGCTGANGCAGTTTTCAGAACACAGG 412  
Db 476 ProfileLeuArgLeuArgAlaGluGlyProGlyThrAlaValArgProProAlaGlnArg 495  
QY 411 GCC-----AGGACCACACACAGGAGGGGGCTTCTATCAAGTTCCCTTCATGCCAT 358  
Db 496 AlaGlyArgArgThrGlyArgArgArgArgGlyAsnHisArg----- 509  
QY 357 ATGAGCCAGNCTGTGCTCGAACCTTGAAGCTGCTGAGAGTCTGGGAC 298  
Db 510 -----ProSerGlyArg 513  
QY 297 NATCCAGGGCTGCATCTTCATCTAGCCTCGTAGCCTTCTGCCTCTCCAGGC 298  
Db 514 ArgProGlyValAlaGlyArgSerSerProAlaArgSerCysArgArgThrArg-gThr 533  
QY 237 TCGTCCATTCTTCAGCTCCAGGCTCCACAGCTCCACAGCTCCATGCCATGCCACCA 178  
Db 534 ArgSerValProSer-----ProThrProSerThrArgTrpAlaLysAlaProArg 550  
QY 177 TGGANACAGCCGTGCCCCCAAGCCAGAGGGAAACAGAACTGGCAAGAAACAGC 118  
Db 551 TrpArgLysProCys-----ThrArgArgSerArgLysAlaGly 563  
QY 117 AGTTAGAGCAGGGGCAACAATGGAGAAAGGAGTATCTATTATGAGCTTGGGCATTTTACA 58  
Db 564 Ser---SerArgLysArgGlyArgThrArgSerMet----- 574  
QY 57 CCAGCACTTCCC 46  
Db 575 ProAlaProPro 578

## RESULT 4

```
US-09-370-838-41
; Sequence 41, Application US/09370838
; Patent No. 644425
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Roadoh
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOUNDS FOR THERAPY AND DIAGNOSIS OF
; TITLE OF INVENTION: LUNG CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.475C1
; CURRENT APPLICATION NUMBER: US/09/370,838
; CURRENT FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/285,323
; EARLIER FILING DATE: 1999-04-02
; NUMBER OF SEQ ID NOS: 289
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 41
; LENGTH: 163
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-370-838-41

Alignment Scores:
Pred. No.: 0.225 Length: 163
Score: 86.00 Matches: 39
Percent Similarity: 43.22% Conservative: 12
Best Local Similarity: 33.05% Mismatches: 50
Query Match: 10.24% Indels: 17
DB: 4 Gaps: 5

US-10-057-510-1 (1-486) x US-09-370-838-41 (1-163)
QY 462 AGGCTTTCGAAACACAGGGGNGTANGNCCAGTTTCAGAACAGAGGGCCAGGACC 403
DB 11 ArgLeuSerLeuArgPro-----SerileProGluLeuSerGluArgThr 25
QY 402 ACAACAGGGGGGGCTTCTATGCAAGTTCCCTTCATCCCATATGAGCCAGAGNCCTN 343
DB 26 SerArgProCysArgAlaSerProAlaSerLeuProSerGlnHisThrSerSerProAla 45
QY 342 GTGGCTCGAACCTTTACTGGGCAATGTCC-----TGAGAGTCTGGGGACNATCCA 292
DB 46 GlnAlaArgValArgAsnLeuAlaGlnSerThrPheProLeuAlaAlaGlnGluThrPro 65
QY 291 GGGGTGATCTTATCTCAGCCCTGCTAGCCTTCTGCTCCTCCTCCAGGGCTCTCC 232
DB 66 GlyArgAlaProAlaHisAlaProLeuSerSerPheValProGlyValGlyArgSer 85
QY 231 ATTCTTCAGCTCCAGCCCTCTCCACAGCTCCATGCTTCCACCTTGCCACATGGA-- 174
DB 86 -----ProAlaSer-ValGlyIleSerAlaProGlyGlyGlyProSerGlyAl 101
QY 173 -NACAGCGTGCCTCC-----ACAAGCCAGGGGAAACAAGAGCTGG 132
DB 101 aAlaAlaIlystleProLeuGluLeuThrGlnSerArgValGlnIlystleTrp 118

RESULT 5
US-09-252-991A-31114
; Sequence 31114, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31114
; LENGTH: 199
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31114

Alignment Scores:
Pred. No.: 0.6 Length: 199
Score: 82.50 Matches: 41
Percent Similarity: 40.91% Conservative: 13
Best Local Similarity: 31.06% Mismatches: 58
Query Match: 9.82% Indels: 20
DB: 4 Gaps: 6

US-09-370-838-41
; SEQ ID NO 31114
; LENGTH: 446
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31114

Alignment Scores:
Pred. No.: 0.316 Length: 446
Score: 86.00 Matches: 43
Percent Similarity: 34.56% Conservative: 4
Best Local Similarity: 31.62% Mismatches: 37
Query Match: 9.82% Indels: 52
DB: 4 Gaps: 9

US-10-057-510-1 (1-486) x US-09-252-991A-31114 (1-446)
QY 29 TGTATTATGGAAGGCTGGGAAGTGTGCTGTGTAATAATGCCAAGCTCATAATAGATCTC 88
DB 65 CysAlaTrpIysArgArgCysSerAlaGlyAla-GlyArg-----77
QY 89 CTTTCTCCATTGTTGCCCTGCTTAACCTGCTGTTCTTTTCCAGCTTCGTTGCC 148
DB 78 -----CysArgCysSerSer-----SerGlyThrProAlaGlyGlyHisPr 91
QY 149 TCTGGCTTGTGGGGCAGCGCTGTNTCCATGTGGCAGGTGGAAGCATGGAGCTGTGGA 208
DB 91 oProAlaSerGlyArgGluAlaArgGly---TrpPro-----102
QY 209 GGAGGGCTGGAGCTGAAGGAATGGACGAGCCCTGGGAGGAGG-----GCAGAAGG 259
DB 103 -----ArgArgTrpProArgProGlyArgArgSerSerAlaAlaGln 117
QY 260 CTAGGAGGCTGTGAG-----ATGAAGATGCAGCCCTCGATNGTCCCA 304
DB 117 yArgCysGlyCysArgGluCysSerGlyAlaGly-ArgCysGluProValGlyPro- 136
QY 305 GACTCTCAGACATTCCTCCAGTCAAGGTTTCGAGCCACNAGNCTTGGCTCATATGGCAT 364
DB 137 -----AlaGlyHisAlaProValArg-----143
QY 365 GAAGGGGAAGCTTCGATA---AGAAGCCCTCTCCTGGTTGTGGT 405
DB 144 --ProGlyArgCysAlaGlyArgAspArgLeuProGlyCysGly 157

RESULT 6
US-09-252-991A-25623
; Sequence 25623, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 25623
; LENGTH: 199
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-25623

Alignment Scores:
Pred. No.: 0.6 Length: 199
Score: 82.50 Matches: 41
Percent Similarity: 40.91% Conservative: 13
Best Local Similarity: 31.06% Mismatches: 58
Query Match: 9.82% Indels: 20
DB: 4 Gaps: 6
```

US-10-057-510-1 (1-486) x US-09-252-991A-25623 (1-199)  
QY 462 AGCTTTGNAACACAGGGGNCCTPANGANCAGTTTCCAGAACAGAGGCCACAGACC 403  
Db 40 ArgThrCysAla---ProAlaGlyGlnArgAlaGlnProGlyGlnArgAspArgPro 58  
QY 402 ACACACAGGAGGGG-----GCTCTTATGCAAGTTCCTTCATGCGCATATG 355  
Db 59 ValGlnArgProGlyGlyArgThrAlaLeuArgArgArgThrGlySerArgHisLeu 78  
QY 354 AGCCAAAGNCTGTGCTCGAACCTTGTAGTGGCAATGCTGTAGAGTCTGGGACNAT 295  
Db 79 AlaGlyAlaAlaCysGlyArgAlaLeuGlnArgAlaLeuSerGluGlyAlaGlyLysPhe 98  
QY 294 CCAGGG-----GCTGCATCTTCATCCTCAGCCCTG 265  
Db 99 ProGlyArgThrLeuAlaGlyAlaGluProProAlaGlyArgGlyValArgValLeu 118  
QY 264 CGTAGCCTTCTGCCCTCCCTCCAGGGC---TCGTCCATTCCTTCA---GCTCCAGCGCT 211  
Db 119 ArgArgHisProAlaPheArgGlyArgProAlaAlaProAspLeuAlaProAlaSer 138  
QY 210 CTTCCACAGTCCATGCTTCCACCTTGCATGTCATGGANACAGCGTGGCCCAACAGCA 151  
Db 139 Pro-----AlaLeuAlaPhe-LeuLeuProArgGlyAlaSerAlaGlyArgSerArgL 156  
QY 150 GAGGGAACACAGAGCTGGCAAGAAAGAACAGCA 117  
Db 156 uArgPheArgArgGlyAlaValGlnLeuProAla 167  
RESULT 7  
US-09-252-991A-24634  
; Sequence 24634, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 24634  
; LENGTH: 320  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-24634  
Alignment Scores:  
Pred. No.: 0.703 Length: 320  
Score: 82.50 Matches: 38  
Percent Similarity: 38.62% Conservative: 18  
Best Local Similarity: 26.21% Mismatches: 69  
Query Match: 9.82% Indels: 20  
DB: Gaps: 6  
US-10-057-510-1 (1-486) x US-09-252-991A-24634 (1-320)  
QY 456 TGCNAAAACACAGGGGNCCTPANGANCAGTTTCCAGAACAGAGGCCACCAACC 397  
Db 12 CysArgArgProSerAla---SerAlaAlaGlyProArgThrProProArgThrSerPhe 30  
QY 396 AGGAGGGGCTCTTATGCAAGTTCCTTCATGCGCATATGAGCAAGNCTGTGGCT 337  
Db 31 SerAlaSerArgProArgValThrSerProThrCysGlyAspCysAlaArgSerThrSer 50  
QY 336 CGAACCTTACTGGGCAATGCTGAGAGTCTGGGACNATCCA-----GGG 289

Db 51 Arg-----TrpArgThrSerSerProAlaProSerProCysAlaSerAlaSer 67  
QY 288 GCTGCATCTTCATCCTCAGCCCTGCTAGCCTTCTGCCCTCTCCACAGGGCTCGTCCATT 229  
Db 68 AlaArgSerGlySerSerArgAlaGlnCysAlaSerThrAlaSerCysSerProIle 87  
QY 228 CCTTCA-----GCTCCAGCGCTCTCCACACGTCATCCATCCCTTCCACCTTGC 181  
Db 88 SerSerProAlaCysArgAlaThrAlaProPro-ArgArgSerSerProGlyThrAlaPr 107  
QY 180 ACATGGNACAGCGTGGCCCAAGCCAGAGGGCAACACGAAGC----- 135  
Db 107 cArgAlaCysThrAlaArgAlaThrSerThrAlaIleArgArgSerAlaCysGlyMetPr 127  
QY 134 -TGCAAGAAAGAACAGCAGTGTAGCAGGCGGCAACAATGGAGAAAGAGTATCTATTAT 76  
Db 127 cTrp-----LysSerAlaMetArgAlaSerSerThrArgArgSerSerCysAla 145  
QY 75 GAGCTTGGGCATT 63  
Db 145 aAsnThrGlyLeu 149  
RESULT 8  
US-09-252-991A-32755  
; Sequence 32755, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 32755  
; LENGTH: 447  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-32755  
Alignment Scores:  
Pred. No.: 0.787 Length: 447  
Score: 82.50 Matches: 40  
Percent Similarity: 37.98% Conservative: 9  
Best Local Similarity: 31.01% Mismatches: 36  
Query Match: 9.82% Indels: 44  
DB: Gaps: 7  
US-10-057-510-1 (1-486) x US-09-252-991A-32755 (1-447)  
QY 401 CAACAGGGAGGGGCTTCTTATGCAAGTTCCTTCATGCCATATGAGCCAGNCCTNG 342  
Db 16 ArgProAlaArgGly-----ProLeuHis-----AlaAlaArgGln 27  
QY 341 TGGGTC-----GAACCTTGTACTGGCAATGTCCTGAGAGTCTGGGACN 297  
Db 28 TrpLeuProGlyLeuArgArgAlaProLeuAlaGlyHisArgProAlaAlaProGlyPro 47  
QY 296 ATCCAGGGGCTGCATCTTCAT----- 276  
Db 48 ProProHisAlaArgArgHisArgGlySerAlaProArgArgGlySerArgThrAlaAla 67  
QY 275 -----CCTCAGCCCTGGTAGCC-----TTCGCGCTCCT 246  
Db 68 SerArgProArgProAlaAlaAlaArgArgProAlaAlaProGlyTrpSerAlaProGly 87  
QY 245 CCCAGGGCTCGTCCATTCTTTCAGTCCAGCGCC---TCCTCCACAGTCCATGCTTCC 189  
Db 88 ProArgAlaAlaProLeuProValArg-AlaGlyAsnAlaValArgProCysArgPr 107



;; PRIOR FILING DATE: 1998-11-09  
;; NUMBER OF SEQ ID NOS: 50  
;; SOFTWARE: PatentIn Ver. 2.1  
;; SEQ ID NO 5  
;; LENGTH: 1917  
;; TYPE: PRT  
;; ORGANISM: Caenorhabditis elegans  
US-09-627-650B-5

Alignment Scores:  
Pred. No.: 1.28 Length: 1917  
Score: 82.50 Matches: 37  
Percent Similarity: 35.71% Conservatives: 8  
Best Local Similarity: 29.37% Mismatches: 40  
Query Match: 9.42% Indels: 41  
DB: Gaps: 6

US-10-057-510-1 (1-486) x US-09-627-650B-5 (1-1917)

QY 99 TGTGCCCCCTGCTCTAAC-----TGCTGTCTCTTCTTCCAGCTTCGTTG 143  
DB 703 CysCysAlaCysThrThrGlyGlyCysAlaAlaCysCysAla----- 716  
QY 144 TTCCTCTGCTGTGGGGCAGCGCTGTTCCATGTGGCAAGTGGAGGCAATGACGT 203  
DB 717 -----CysAlaCysAlaThrAlaAlaCysThr-CysGlyThrThr-CysCysThrThrC 734  
QY 204 GT-----GGAGGAGCGCTGGAGCTGAAGGATGGAAGTGGAGCGCCCT 242  
DB 734 ySgLyThrAlaThrCysGlyAlaGlyGlyGlyThrGlyAlaThrGly--GlyAlaAlaCy 753  
QY 243 GGGAGGAGGCGAGAGCTACGACGCGCTGAGGATGAAGATGAGATGCGATGCGTCC 302  
DB 753 sGlyGlyThrThrAlaThr-----AlaCysThr----- 763  
QY 303 CAGACTCTCAGGACATTGCCAGTCAAGGGTTCGAGCCACNAGNCTTGGCTCATATGGC 362  
DB 764 -----AlaGlyThrCysAlaAlaAlaGlyAlaThrThrAlaAlaCysAla----- 778  
QY 363 ATGAAGGGAACTTGCATAAGAGCCCTCCCTGCTGTGTGGTCTGCTGCTGCTGCTG 422  
DB 779 -----GlyThrCysAlaCysThrGlyCysAlaAlaCysGlyThrGlyThrCysCysAl 796  
QY 423 GAAACTGCGN 432  
DB 796 aAlaThrGly 799

RESULT 12

US-09-436-063C-5  
; Sequence 5, Application US/09436063C  
; Patent No. 6407210  
; GENERAL INFORMATION:  
; APPLICANT: Bamber, Bruce  
; APPLICANT: Jorgensen, Erik  
; TITLE OF INVENTION: Nematode Neuromuscular Junction GABA Receptors and  
; FILE REFERENCE: P-1095corrected  
; CURRENT APPLICATION NUMBER: US/09/436,063C  
; PRIOR FILING DATE: 1999-11-08  
; PRIOR APPLICATION NUMBER: 60/107727  
; PRIOR FILING DATE: 1998-11-09  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 1917  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans  
US-09-436-063C-5

Alignment Scores:  
Pred. No.: 1.28 Length: 1917  
Score: 82.50 Matches: 37  
Percent Similarity: 35.71% Conservatives: 8

Best Local Similarity: 29.37% Mismatches: 40  
Query Match: 9.42% Indels: 41  
DB: Gaps: 6

US-10-057-510-1 (1-486) x US-09-436-063C-5 (1-1917)

QY 99 TGTGCCCCCTGCTCTAAC-----TGCTGTCTCTTCTTCCAGCTTCGTTG 143  
DB 703 CysCysAlaCysThrThrGlyGlyCysAlaAlaCysCysAla----- 716  
QY 144 TTCCTCTGCTGTGGGGCAGCGCTGTTCCATGTGGCAAGTGGAGGCAATGACGT 203  
DB 717 -----CysAlaCysAlaThrAlaAlaCysThr-CysGlyThrThr-CysCysThrThrC 734  
QY 204 GT-----GGAGGAGCGCTGGAGCTGAAGGATGGAAGTGGAGCGCCCT 242  
DB 734 ySgLyThrAlaThrCysGlyAlaGlyGlyGlyThrGlyAlaThrGly--GlyAlaAlaCy 753  
QY 243 GGGAGGAGGCGAGAGCTACGACGCGCTGAGGATGAAGATGAGATGCGATGCGTCC 302  
DB 753 sGlyGlyThrThrAlaThr-----AlaCysThr----- 763  
QY 303 CAGACTCTCAGGACATTGCCAGTCAAGGGTTCGAGCCACNAGNCTTGGCTCATATGGC 362  
DB 764 -----AlaGlyThrCysAlaAlaAlaGlyAlaThrThrAlaAlaCysAla----- 778  
QY 363 ATGAAGGGAACTTGCATAAGAGCCCTCCCTGCTGTGTGGTCTGCTGCTGCTGCTG 422  
DB 779 -----GlyThrCysAlaCysThrGlyCysAlaAlaCysGlyThrGlyThrCysCysAl 796  
QY 423 GAAACTGCGN 432  
DB 796 aAlaThrGly 799

RESULT 13

US-09-627-650B-7  
; Sequence 7, Application US/09627650B  
; Patent No. 6406872  
; GENERAL INFORMATION:  
; APPLICANT: Bamber, Bruce  
; APPLICANT: Jorgensen, Erik  
; TITLE OF INVENTION: Nematode Neuromuscular Junction GABA Receptors and  
; FILE REFERENCE: 21101.0009U3  
; CURRENT APPLICATION NUMBER: US/09/627,650B  
; CURRENT FILING DATE: 2000-07-28  
; PRIOR APPLICATION NUMBER: 09/436,063  
; PRIOR FILING DATE: 1999-11-08  
; PRIOR APPLICATION NUMBER: 60/107,727  
; PRIOR FILING DATE: 1998-11-09  
; NUMBER OF SEQ ID NOS: 50  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 7  
; LENGTH: 2508  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans  
US-09-627-650B-7

Alignment Scores:  
Pred. No.: 1.4 Length: 2508  
Score: 82.50 Matches: 37  
Percent Similarity: 35.71% Conservatives: 8  
Best Local Similarity: 29.37% Mismatches: 40  
Query Match: 9.42% Indels: 41  
DB: Gaps: 6

US-10-057-510-1 (1-486) x US-09-627-650B-7 (1-2508)

QY 99 TGTGCCCCCTGCTCTAAC-----TGCTGTCTCTTCTTCCAGCTTCGTTG 143  
DB 543 CysCysAlaCysThrThrGlyGlyCysAlaAlaCysCysAla----- 556  
QY 144 TTCCTCTGCTGTGGGGCAGCGCTGTTCCATGTGGCAAGTGGAGGCAATGACGT 203





Wed Apr 7 07:32:31 2004

Search completed: April 6, 2004, 17:59:23  
Job time : 25.5 secs

us-10-057-510-1.ra1

Page 8

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: April 6, 2004, 17:48:14 ; Search time 54.5 Seconds  
(without alignments)  
5039.196 Million cell updates/sec

Title: US-10-057-510-1

Perfect score: 876  
Sequence: 1 gcacgagcgtatnaaaat.....aaagggaantccccctggg 486

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 3172214

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp  
-O=/cn2.1/USPTO.spool.p/US10057510/runat\_06042004\_141608\_14698/app.query.fasta\_1.647  
-DB=A\_Geneseq\_29Jan04 -QFMT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=ptc -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -CUTFM=ptc -NORW=ext -HEAFSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10057510 @CGN\_1.1.81/runat\_06042004\_141608\_14698 -NCFU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A\_Geneseq\_29Jan04:\*

1: Geneseqp1980s: \*  
2: Geneseqp1990s: \*  
3: Geneseqp2000s: \*  
4: Geneseqp2001s: \*  
5: Geneseqp2002s: \*  
6: Geneseqp2003as: \*  
7: Geneseqp2003bs: \*  
8: Geneseqp2004s: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	394	45.0	74	3	Aab08845 A human M
2	387	44.2	247	2	Aay48595 Human bre
3	183.5	20.9	196	5	Abb90052 Human pol
C 4	93.5	11.1	575	4	Abg21970 Novel hum
5	92.5	10.6	288	4	Abg04148 Novel hum
C 6	91	10.8	1373	4	Abb62368 Drosophil
C 7	89.5	10.7	647	2	Aaw04327 Rat petri
C 8	89.5	10.7	1708	5	Abp69375 Human pol
C 9	89	10.2	239	5	Abg60094 Human DIT
C 10	88.5	10.5	247	6	Aae35942 Lolium pe

C 11	86	10.2	163	2	AAY29526	Aay29526 Human lun
C 12	86	10.2	163	3	AAB44416	Aab44416 Human lun
C 13	86	10.2	163	4	AAE13757	Aae13757 Human lun
C 14	86	10.2	163	7	ADD66349	Add66349 Human lun
C 15	86	10.2	163	7	AD887603	Ad887603 Human lun
16	84	9.6	338	2	AAR73955	Aar73955 Duffy blo
17	84	9.6	338	2	AAW54857	Aaw54857 Homo sapi
18	84	9.6	338	5	ABB76196	Abb76196 Human Duf
19	84	9.6	338	6	ABP81813	Abp81813 Human Duf
20	84	9.6	338	6	ABR92081	AbR92081 Human cer
21	83.5	9.9	266	4	ABB70975	Abb70975 Drosophil
C 22	83.5	9.9	817	6	ABR53281	AbR53281 Protein s
C 23	83.5	9.9	3502	4	ABBS8382	Abbs8382 Drosophil
C 24	83	9.9	202	4	AAW79674	Aam79674 Human pro
C 25	83	9.9	213	4	AAW78690	Aam78690 Human pro
C 26	83	9.9	221	4	AAAB29791	Aab29791 Human ste
C 27	83	9.9	239	3	AAAB33294	Aab33294 Human col
C 28	83	9.9	239	4	AAG75034	Aag75034 Human col
C 29	83	9.9	426	4	AAG62628	Aag62628 Chlorella
C 30	83	9.5	542	4	ABB11016	Abb11016 Human sec
31	83	9.5	2087	4	AAAB31516	Aab31516 Amino aci
32	83	9.5	2091	3	AAAB12000	Aab12000 Rat p3103
33	82.5	9.4	213	4	AAU19618	Aau19618 Human dia
C 34	82.5	9.8	511	5	ABG93255	Abg93255 C. albica
35	82.5	9.4	1652	6	ADA15715	Ada15715 C. elegan
36	82.5	9.4	1917	6	ADA15719	Ada15719 C. elegan
37	82.5	9.4	2508	6	ADA15721	Ada15721 C. elegan
38	82.5	9.4	2544	6	ADA15717	Ada15717 C. elegan
39	82.5	9.4	2601	6	ADA15723	Ada15723 C. elegan
C 40	82	9.8	363	4	ABB68590	Abb68590 Drosophil
C 41	82	9.8	1386	7	ADES9291	AdeS9291 Rat Prote
C 42	82	9.8	1386	7	ADES9295	AdeS9295 Rat Prote
C 43	82	9.8	19938	6	ABP76682	Abp76682 Streptomy
C 44	81.5	9.7	560	4	ABG21040	Abg21040 Novel hum
C 45	81.5	9.3	963	4	ABG02396	Abg02396 Novel hum

ALIGNMENTS

RESULT 1

AAB08845

ID AAE08845 standard; protein; 74 AA.

XX AC AAE08845;

DT 02-JAN-2001 (first entry)

XX DE A human MDM2 interacting polypeptide (MDMIP).

XX KW Human; MDM2 interacting polypeptide; MDMIP; MDM2; cell cycle progression; cell differentiation; cancer; glioma; squamous cell carcinoma; breast cancer; astrocytoma; leukemia; lymphoma; tumorigenesis; gene therapy.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

FT Misc-difference 5 /note= "unknown amino acid encoded by NAA"

FT FT Misc-difference 58

FT FT Misc-difference 58 /note= "unknown amino acid encoded by TNT"

XX PN WO200050590-A1.

XX PD 31-AUG-2000.

XX PF 23-FEB-2000; 2000WO-US004582.

XX PR 23-FEB-1999; 99US-0121192P.

XX PR 03-MAR-1999; 99US-0122643P.

XX PR 22-FEB-2000; 2000US-00510252.

XX PA (CURA-) CURAGEN CORP.

XX PI Nandabalan K, Yang M, Schulz VP;  
 XX DR WPI: 2000-558398/51.  
 XX DR N-PSDB; AA75041.  
 XX FT Novel MDM2 interacting protein useful for treating or preventing  
 XX FT disorders involving aberrant levels of MDM2 and/or MDMV-interacting  
 XX FT proteins, comprises a specific amino acid sequence.  
 XX PS Claim 10; Fig 1; 78pp; English.  
 XX CC The present sequence represents a human MDM2 interacting polypeptide  
 XX CC (MDMIP). MDMIP was identified using a yeast two hybrid system, using a  
 XX CC fragment of MDM2 as the bait protein. The MDMIP polypeptide is useful for  
 XX CC detecting and removing MDM2 polypeptides in a biological sample by  
 XX CC forming MDM2-MDMIP complexes. MDMIP and MDM2 are useful to identify  
 XX CC compounds or other agents which modulate the activity of MDM2 and/or  
 XX CC MDMIP-mediated processes. Agents that modulate the function of MDMIP/MDM2  
 XX CC complexes are useful for treating and preventing a disease or disorder  
 XX CC involving aberrant levels of MDM2 or MDMIP. MDMIP is also useful for  
 XX CC treating diseases caused by aberrant levels of expression of MDM2 genes,  
 XX CC such as disorders of cell cycle progression, cell differentiation, and  
 XX CC transcriptional control, including cancers such as human sarcoma, glioma,  
 XX CC squamous cell carcinoma, breast cancer, astrocytoma, leukemia and  
 XX CC lymphoma, and tumorigenesis. MDMIP and MDM2 nucleic acids are useful in  
 XX CC gene therapy  
 XX SQ Sequence 74 AA;  
 Alignment Scores:  
 Pred. No.: 2,01e-32 Length: 74  
 Score: 394.00 Matches: 74  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 44.98% Indels: 0  
 DB: 3 Gaps: 0  
 US-10-057-510-1 (1-486) x AAB08845 (1-74)  
 QY 1 GCACGAGCGTATNAAAAATTTTCCATGTTTATGAAAGCGTGGAGTCTGGTGT 60  
 Db 1 AlaArgAlaTyr\*\*\*LysMetPheSerMetPheMetGluLysAlaGlyLysCysTrpCys 20  
 QY 61 AAATGCCCAAGCTCATATAGACTCCCTTCTCCATTTCTGCCCTGCTCTAATGCT 120  
 Db 21 LysMetProLysLeuIlelleAspThrProPheSerIleValAlaProAlaLeuThrAla 40  
 QY 121 GTTCTTTCTTCCAGCTTCGTTGTTTCCCTCTGCTTGGGGCGCAGGCTGNTCCATGT 180  
 Db 41 ValLeuSerCysGlnLeuArgCysSerLeuTrpLeuValGlyAlaArgLeu\*\*\*ProCys 60  
 QY 181 GGCAGGTGGAGCATGGACGTGTGGAGGAGCGCTGGAGC 222  
 Db 61 GlyLysValGluGlyMetAspValTrpArgArgTrpSer 74  
 RESULT 2  
 AAY48585  
 ID AAY48585 standard; protein, 247 AA.  
 XX AC AAY48585;  
 XX XX  
 DT 08-DEC-1999 (first entry)  
 XX DE Human breast tumour-associated protein 46.  
 XX KW Expressed sequence tag; EST; human; breast; cancer; gene therapy;  
 XX KW treatment; tumour; cytostatic; medicament.  
 XX OS Homo sapiens.  
 XX XX  
 FN DE19813839-A1.  
 XX XX

PD 23-SEP-1999.  
 XX XX  
 PF 20-MAR-1998; 98DE-01013839.  
 XX XX  
 PR 20-MAR-1998; 98DE-01013839.  
 XX XX  
 PA (META-) METAGEN GES GENOMFORSCHUNG MBH.  
 XX XX  
 PI Specht T, Hinzmann B, Schmitt A, Pillarsky C, Dahl E, Rosentahl A;  
 XX XX  
 DR WPI: 1999-528981/45.  
 DR N-PSDB; AA233655.  
 XX XX  
 PT Human nucleic acid sequences and protein products from tumor breast  
 XX tissue, useful for breast cancer therapy.  
 XX PS Claim 22; 163; 188pp; German.  
 XX CC This invention describes novel human nucleic acid sequences from tumor  
 XX CC breast tissue which have cytostatic activity. The nucleic acid sequences  
 XX CC can be used to produce and isolate full-length gene sequences. They can  
 XX CC be used to express proteins, which can be used as tools to find an  
 XX CC activity against breast cancer. The sequences can be used in sense or  
 XX CC antisense form. They are especially useful for medicaments for gene  
 XX CC therapy to treat breast cancer. AA48540-Y48617 represent protein  
 XX CC fragments encoded by the expressed sequence tags described in the method  
 XX CC of the invention  
 XX SQ Sequence 247 AA;  
 Alignment Scores:  
 Pred. No.: 1.4e-31 Length: 247  
 Score: 387.00 Matches: 94  
 Percent Similarity: 72.99% Conservative: 6  
 Best Local Similarity: 68.61% Mismatches: 30  
 Query Match: 44.18% Indels: 8  
 DB: 2 Gaps: 2  
 US-10-057-510-1 (1-486) x AAY48585 (1-247)  
 QY 87 TCCTTTCTCCATGTTGGCCCTGCTCTAACTGCTGTTCTTTCTTCCAGCTTCGTTGTTTC 146  
 Db 34 SerPheLeuHisAlaThrHisHisLeuSer---AlaGluValLysAlaAlaSerLeuPhe 52  
 QY 147 CCTCTGCTTGTGGGGCGCAGGCTGTNTC-CATGTGGCAAGGTGAAGCATGACGTGT 205  
 Db 53 ProLeuAlaCysGlyGlyThrAlaValLeuHisValAlaAargTrpLysAlaTrpThrCys 72  
 QY 206 GGAGGAGCGCTGAGCTGAAGGAATGGACGAGCCCTGGAGGAGGAGGAGGAGGAGGAG 265  
 Db 73 GlyGlyGlyAlaGlyAlaGlyAlaGlyMetAspGluProTrpGluGluGlyArgLeuArg 92  
 QY 266 AGGCTGAGGATGAAGATGAGCCCTGATGATGTCCTCCAGACTCTCAGGACATTCGCCAG 325  
 Db 93 ArgAlaGluAspGluAspAlaAlaProGlyTrpSer-GlnThrLeuArgThrCysProAl 112  
 QY 326 TCAAGGTTTCGAGCCACNAGNCTTGGCTCATATGGCATGAAGGGAGGAGGAGGAGGAG 385  
 Db 112 aGlnGlyLeuArgAlaThrGlyLeuAlaSerTyrglyMetArgGlySerTrpHis-ArgS 132  
 QY 386 GCCCCTCCCTGTTGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 445  
 Db 132 erProLeuPro-AlaValValLeu-ProSerValLeuGlnThr---AlaLeuSerProbe 150  
 QY 446 GGTTTTNGCAAGCCTGAAGAAAGGAAANTCCCATGGG 486  
 Db 150 uAlaLeuCysGlnAlaTrpArg-ArgAlaValProHisGly 163  
 RESULT 3  
 ABB90052  
 ID ABB90052 standard; protein, 196 AA.  
 XX XX  
 AC ABB90052;

XX 24-MAY-2002 (first entry)  
 XX Human polypeptide SEQ ID NO 2428.  
 XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
 KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;  
 KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
 KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
 KW neurological disease; infection; human; secreted protein.  
 XX Homo sapiens.  
 XX WO200190304-A2..  
 XX 29-NOV-2001.  
 XX 18-MAY-2001; 2001WO-US016450.  
 XX 19-MAY-2000; 2000US-0205515P.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX Birse CE, Rosen CA;  
 XX WPI; 2002-122018/16.  
 XX N-PSDB; ABL90461.  
 XX Novel 1405 isolated polypeptides, useful for diagnosis, treatment and  
 PT prevention of neural, immune system, muscular, reproductive, and  
 PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative  
 PT disorders.  
 XX Claim 11; SEQ ID NO 2428; 2081pp + Sequence Listing; English.  
 XX The invention relates to novel genes (ABL89449-ABL90853) and proteins  
 CC (AB89040-AB93044) useful for preventing, treating or ameliorating  
 CC medical conditions e.g. by protein or gene therapy. The genes are  
 CC isolated from a range of human tissues disclosed in the specification.  
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in  
 CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and  
 CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,  
 CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune  
 CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic  
 CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,  
 CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)  
 CC cardiovascular disorders such as myocardial ischaemia; (d) wound healing  
 CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)  
 CC infectious diseases such as viral, bacterial, fungal and parasitic  
 CC infections. Note: The sequence data for this patent did not form part of  
 CC the printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX Sequence 196 AA;  
 XX  
 XX Alignment Scores:  
 XX Pred. No.: 2,518-10 Length: 196  
 XX Score: 183.50 Matches: 51  
 XX Percent Similarity: 69.14% Conservative: 5  
 XX Best Local Similarity: 62.96% Mismatches: 19  
 XX Query Match: 20.95% Indels: 6  
 XX DB: 5 Gaps: 1  
 XX  
 XX US-10-057-510-1 (1-486) x ABB90052 (1-196)  
 QY 254 AGAAGCTACGAGGCTGAGTGAAGATGAGCCCTGGTGTGCTCCGACTCTCAG 313  
 DB 38 ArgArgLeuArgAlaGluAspGluAspAlaAlaProGlyTrpSer-GlnThrLeuAr 57  
 QY 314 GACATGCCAGTCAAGGGTTCAGGCCACNAGNCTTGGCTCATATGGCATGAGGGGAA 373  
 DB 57 gthCysProAlaGlnGlyLeuArgAlaThrGlyLeuAlaSerTyrGlyMetArgGlySe 77

QY 374 CTTGCATAAGAGCCCTCCTGCTGTGTGGTCTGCTGCGCCCTGTGTGGAACCTGGNT 433  
 DB 77 rTrpHis-ArgSerProLeuPro-AlaValValLeu-ProSerValLeuGlnThr---Al 95  
 QY 434 CNTAGNCCCTGGTGTGTTTNGCAAGCCTGAAGAAAGGAAANTCCCATGGG 486  
 DB 95 aleuSerProLeuAlaLeuCysGlnAlaTrpArg-ArgAlaValProHisGly 112  
 RESULT 4  
 ABG21970  
 ID ABG21970 standard; protein; 575 AA.  
 XX  
 XX ABG21970;  
 XX DT 18-FEB-2002 (first entry)  
 XX DE Novel human diagnostic protein #21961.  
 XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 XX KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX OS Homo sapiens.  
 XX PN WO200175067-A2.  
 XX PD 11-OCT-2001.  
 XX PF 30-MAR-2001; 2001WO-US008631.  
 XX PR 31-MAR-2000; 2000US-00540217.  
 XX BR 23-AUG-2000; 2000US-00649167.  
 XX PA (HYSE-) HYSEQ INC.  
 XX  
 XX Drmanac RT, Liu C, Tang YT;  
 XX WPI; 2001-639362/73.  
 XX N-PSDB; AAS86157.  
 XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity.  
 XX Claim 20; SEQ ID NO 52329; 103pp; English.  
 XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 CC sequences. (I) is useful as hybridisation probes, polymerase chain  
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 CC and in recombinant production of (II). The polynucleotides are also used  
 CC in diagnostics as expressed sequence tags for identifying expressed  
 CC genes. (I) is useful in gene therapy techniques to restore normal  
 CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a  
 CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical imaging  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological activity. The  
 CC polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic  
 CC amino acid sequences of the invention. Note: The sequence data for this  
 CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 XX Sequence 575 AA;  
 XX  
 XX Alignment Scores:  
 XX Pred. No.: 0.809 Length: 575  
 XX Score: 93.50 Matches: 36

Percent Similarity: 40.32% Conservative: 14  
Best Local Similarity: 29.03% Mismatches: 57  
Query Match: 11.13% Indels: 17  
DB: 4 Gaps: 3

US-10-057-510-1 (1-486) x ABG21970 (1-575)

```
QY 447 CCAGGGGNCNTANGANCAGTTTCCAGAACAGAGGGCCAGG---ACCACACACAGGAG 391
DB 198 ProGlyHisGlnArgSerLeuMetProProSerArg***ArgAlaThrSerThrAlaTrp 217
QY 390 GGGGCTTCTTATGCAAGTCCCTTCATGCCATATGAGCCAGACCTNGTGGCTCGAAC 331
DB 218 ArgAlaArgThrAlaSerThrProArgCys-----LeuGlyProArgPro 232
QY 330 CTGTGCTGGGCAATCTCTGAGAGTCTGGGACNATCCAGGGGCTGCATCTCATCTCA 271
DB 233 AlaProTrpSerCys-----AlaSerLeuSerSerSer 243
QY 270 GCCTGCGTAGCTCTGCTCCCTCCAGGGCTGCTCCATCTCTCAGCTCCAGCGCT 211
DB 244 HisArgProAsnThrGlnProSerProSerProThrSerValArgArgAlaArgLeuAla 263
QY 210 CTCCACACGTCCTCCCTTCCACCTTGCACATGGANACAGCCGCTGCCCCCAAGCCA 151
DB 264 ThrProSerSerHisGlyTrpAlaLeuAlaProTrpThrProMetAlaProThrGlnPro 283
QY 150 GAGGACACAGAGCTGCGCAAGAACAGACAGCAGTGTAGACAGCGGGCAACATGGAGAA 91
DB 284 ThrThrTrpThrArg-TrpGlyLysThrValSerGlyArg-GlnMetAsnThrTrpArg 303
QY 90 AGGAGTAT 83
DB 303 rgProTrp 305
```

RESULT 5  
ID ABG04148  
ABG04148 standard; protein; 288 AA.

XX AC ABG04148;  
XX DT 13-FEB-2002 (first entry)  
XX DE Novel human diagnostic protein #4139.  
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX KM food supplement; medical imaging; diagnostic; genetic disorder.  
XX OS Homo sapiens.

XX WO200175067-A2.  
XX PD 11-OCT-2001.  
XX PF 30-MAR-2001; 2001WO-US008631.  
XX PR 31-MAR-2000; 2000US-00540217.  
XX PR 23-AUG-2000; 2000US-00649167.  
XX PA (HYSE-) HYSEQ INC.  
XX PI Drmanac RT, Liu C, Tang YT;  
XX DR WPI; 2001-639362/73.  
XX DR N-PSDB; AAS68335.

XX New isolated polynucleotide and encoded polypeptides, useful in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits and to assess  
XX biodiversity.

XX Claim 20; SEQ ID NO 34507; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
CC sequences. (I) is useful as hybridisation probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping.  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activities in  
CC polypeptide and polynucleotide sequences have application in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG0010-ABG3037 represent novel human diagnostic  
CC amino acid sequences of the invention. Note: the sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 288 AA;

Alignment Scores:  
Pred. No.: 0.887 Length: 288  
Score: 92.50 Matches: 37  
Percent Similarity: 37.67% Conservative: 18  
Best Local Similarity: 25.34% Mismatches: 50  
Query Match: 10.56% Indels: 41  
DB: 7 Gaps: 7

US-10-057-510-1 (1-486) x ABG04148 (1-288)

```
QY 40 AAGCTGGGAAGTGTGCTGTAATAATGCCAAGCTCATATAGATACTCTTCTCCATT 99
DB 48 LysSerHisGluLeuPheCysLysMetGluProGlnValLeuGluProArgAlaVal 67
QY 100 GTT-----GCCCTGCTCTAACTGCTGTCTTCTTCTGCGAGCTTGTGTTCCCTC 150
DB 68 LeuGlnAspGlyAlaProGlyProGlyArgAlaThrSerCysSerAlaArg----- 84
QY 151 TGGCTTGTGGGGCAGCGCTGTNTCCATGTGGCAAGTGTGCAAGGCATGACGCTGTGGAGG 210
DB 85 -----LysGlyArgGlyProGluLysProValGlnGlyLeuPro----- 97
QY 211 AGCGCTGGAGCTGAAGAAATGGA-----CGAGCCCTGGGAGGAGGAGAGGCTACG 264
DB 98 -----AsnGlySerValArgAlaHisSerGlyGlyArgAlaAlaPro 111
QY 265 CAGGCTGAGGATGAAGATGCAGCCCTGGATNGTCCCGAGCTCTCAGACATTTGCCCA 324
DB 112 Gln-----ProSerProArgGlyHisGlyPro 120
QY 325 GTCAAGGGTTTCAGCCACNAGNCTTGGCTCATATGCGCATGAAGGGGAACCTGCATAAGA 384
DB 121 GlyArgGly***ThrAlaAlaProLeuProHisLeu-----CysProLeu 135
QY 385 AGCCCTCTCCCTGGTGTGCTGCTGCGCCCTCTGTCTGTGAAAACTGGTCTAGTACCC 444
DB 136 ThrProValLeuLeuGlnGly***GlyProAsp***TrpProLeuGly----- 151
QY 445 TGGTTTTTNGCAAGCCT 462
DB 152 TrpAlaThrMetArgPro 157
```

RESULT 6  
ID ABB62368  
ABB62368 standard; protein; 1373 AA.  
XX ABB62368;  
XX DT 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 13896.  
 DE Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.  
 KW Drosophila melanogaster.  
 OS WO200171042-A2.  
 FN 27-SEP-2001.  
 PD 23-MAR-2001; 2001WO-US009231.  
 XX 23-MAR-2000; 2000US-0191637P.  
 PR 11-JUL-2000; 2000US-00614150.  
 FR (PEKE) PE CORP NY.  
 XX Venter JC, Adams M, Li PWD, Myers EW;  
 PI WPI; 2001-656860/75.  
 DR N-PSDB; ABL06471.

XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signaling and cell-cell  
 PT interactions.  
 XX Disclosure; SEQ ID NO 13896; 21bp + Sequence Listing; English.  
 XX The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-  
 CC ABB72072). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX Sequence 1373 AA;

Alignment Scores:  
 Pred. No.: 1.78 Length: 1373  
 Score: 91.00 Matches: 46  
 Percent Similarity: 41.40% Conservative: 19  
 Best Local Similarity: 29.30% Mismatches: 63  
 Query Match: 10.83% Indels: 29  
 Gaps: 7

US-10-057-510-1 (1-486) x ABB62368 (1-1373)  
 QY 441 GGNCTANGANCAGTTTTCAGACAGAGGCCAGGACCAACACAGGAGGGGCTTCT 382  
 Db 1029 GlyValValAenThrAsnProValArgProArgSerAlaSerSerHisGlyHisSer 1048  
 QY 361 TATGCAAGT-----TCCCTTCATGCCATATGAGCCAGNCCTNGTGCT 337  
 Db 1049 ValGlySerThrSerAlaProThrSerProSerValHisAlaSerSerGlnValIleLys 1068  
 QY 336 CGAACCCCTTGAGTGGCCCAATGCTCTGAGAGTCTGGGAC-----NATCCAGGG 289  
 Db 1069 ArgSerSerProAlaArgSerGlnAlaSerGlnGlyAlaLeuAspLeuThrProArg 1088  
 QY 288 GTGCA---TCCTTATCTCAGCCCTGGTAGCCCTTCTGCCCTCTCCAGGGGCTCTCC 232  
 Db 1089 AlaAlaProThrSerSerSerSerArgSerProLeuProLysGluLysProValSer 1108  
 QY 231 ATTCTCTAGCTCCAGCGC-----CTCCTCCA 205  
 Db 1109 ProProSerLeuPro-ArgSerProSerGlySerSerHisAlaSerAlaAsnIleLeuTh 1128

QY 204 CACGTCCATGCTTCCACC-----TTGCCACATGGANACAGCCGTGCCCC 160  
 Db 1128 rSerProLeuProProThrValGlyLeuAspCysLeuProGlyLeuGlnHis---Hi 1147  
 QY 159 CACAGCCAGAGGGAACACAGAGCTGGCAAGAAAGACAGCAGTTAGAGCAGGGGCAAC 100  
 Db 1147 sLeuGlnGlnHisGlnHisLeuMetGlnGlnGlnAlaAlaAlaAlaAlaAla 1167  
 QY 99 AATGGAGAAAGAGTATCTATTATGAGCTTGGGCTTTTACACCAGCAC 51  
 Db 1167 aalaglnHisHis---HisHisGlnGlnMetAlaAlaLeuHisGlnHis 1182

## RESULT 7

AAW04327  
 ID AAW04327 standard; protein; 647 AA.

XX AC AAW04327;

XX DT 16-JAN-1997 (first entry)

XX DE Rat petrin.

XX KW Petrin; neurite outgrowth associated protein; CNS;  
 KW central nervous system; myelin; protein phosphatase 2C; stroke;  
 KW neurodegeneration.

XX OS Rattus sp.

XX PN WO9632476-A1.

XX PD 17-OCT-1996.

XX PF 12-APR-1996; 96WO-CA000214.

XX PR 13-APR-1995; 95US-00421701.

XX PA (MOUN) MOUNT SINAI HOSPITAL CORP.

XX PI Roach A, Lozano A, Labes M, Roder J;

XX DR WPI; 1996-477127/47.

XX DR N-PSDB; AAT38486.

XX Assay for substance that modulates response of neuronal cells - and  
 PT neurite growth associated protein, Petrin, useful in conditions involving  
 PT nerve damage resulting from traumatic injury, stroke or CNS degenerative  
 PT disorders.

XX PS Claim 10; Fig 24; 119pp; English.

XX CC Rat petrin (AAW04327) is a protein involved in modulating neurite growth  
 CC inhibition. Its amino sequence was deduced from a cDNA sequence  
 CC (AAT38486) derived from an adult rat brain cDNA library. Petrin is a new  
 CC member of the protein phosphatase 2C family, and is expressed in neurons  
 CC in brain tissue, partic. in the Purkinje cells of the cerebellum. Petrin,  
 CC and antibodies raised against it, can be used to modulate neurite growth  
 CC and axonal regeneration

XX SQ Sequence 647 AA;

## Alignment Scores:

Pred. No.: 2.17 Length: 647  
 Score: 89.50 Matches: 39  
 Percent Similarity: 41.60% Conservative: 13  
 Best Local Similarity: 31.20% Mismatches: 52  
 Query Match: 10.65% Indels: 21  
 Gaps: 6

US-10-057-510-1 (1-486) x AAW04327 (1-647)

QY 447 CCAGGGGGNCTANGANCAGTTTTCAGAACAG---AGGCCAGGACCAACACAGGAG 391

Db 17 ProGlyGlyLeuSerSerAlaHisAlaGluThrGlyArgGlyAlaGlyLysArgPro 36



KW gastrointestinal disorder; transport disorder; Gene therapy; kidney;  
KW adrenal gland; bone marrow; lung; ovary; pancreas; prostate; spleen;  
KW thymus.  
XX Homo sapiens.  
XX OS  
XX WO200220754-A2.  
XX PD 14-MAR-2002.  
XX 29-AUG-2001; 2001WO-US027127.  
XX 05-SEP-2000; 2000US-0229747P.  
XX PR 05-SEP-2000; 2000US-0229748P.  
XX PR 05-SEP-2000; 2000US-0229749P.  
XX PR 05-SEP-2000; 2000US-0229750P.  
XX PR 05-SEP-2000; 2000US-0229751P.  
XX PR 05-SEP-2000; 2000US-0230583P.  
XX PR 06-SEP-2000; 2000US-0230585P.  
XX PR 06-SEP-2000; 2000US-0230514P.  
XX PR 06-SEP-2000; 2000US-0230515P.  
XX PR 06-SEP-2000; 2000US-0230517P.  
XX PR 06-SEP-2000; 2000US-0230518P.  
XX PR 06-SEP-2000; 2000US-0230519P.  
XX PR 06-SEP-2000; 2000US-0230593P.  
XX PR 06-SEP-2000; 2000US-0230597P.  
XX PR 06-SEP-2000; 2000US-0230598P.  
XX PR 06-SEP-2000; 2000US-0230599P.  
XX PR 06-SEP-2000; 2000US-0230610P.  
XX PR 06-SEP-2000; 2000US-0230865P.  
XX PR 06-SEP-2000; 2000US-0230888P.  
XX PR 07-SEP-2000; 2000US-0230951P.  
XX PR 07-SEP-2000; 2000US-0231163P.  
XX PR 07-SEP-2000; 2000US-0231167P.  
XX (INCY-) INCYTE GENOMICS INC.  
XX Stuart J, Lincoln SE, Altus CM, Dufour GE, Chalup MS, Hillman JL;  
PI Jones AU, Yu JY, Wright RJ, Gietzen D, Liu TF, Yap PE, Dahl CR;  
PI Momiya MC, Bradley DL, Rohatgi SD, Harris B, Roseberry AW;  
PI Gerstin EH, Peralta CH, David MH, Panzer SR, Flores V, Daiffo A;  
PI Marwaha R, Chen AJ, Chang SC, Au AP, Inman RR;  
XX WPI: 2002-383054/41.  
DR N-PSDB; ABK71875.  
XX An isolated polynucleotide useful in diagnostics and therapeutics.  
XX  
XX  
XX Claim 29; Page 614-615; 686pp; English.  
XX The invention relates to human diagnostic and therapeutic (dithp)  
CC polynucleotides and their associated polypeptides (DITHP polypeptides).  
CC The sequences of the invention are used in the treatment and diagnosis of  
CC cell proliferative disorders (e.g. atherosclerosis, cirrhosis), cancers  
CC (e.g. tumours of the adrenal gland, bone, bone marrow, brain, breast,  
CC cervix, kidney, lung, ovary, pancreas, prostate, skin, spleen, testis or  
CC thymus), autoimmune/inflammatory disorders (e.g. asthma, bronchitis,  
CC psoriasis, osteoporosis), viral infections, bacterial infections, fungal  
CC infections, parasitic infections, developmental disorders (e.g. anaemia,  
CC epilepsy), seizure disorders (e.g. cerebral palsy, spina bifida),  
CC endocrine disorders (e.g. thrombosis, aneurysm), metabolic disorders  
CC (e.g. obesity, diabetes), neurological disorders (e.g. stroke,  
CC amyotrophic lateral sclerosis, multiple sclerosis), gastrointestinal  
CC disorders (e.g. ulcerative colitis, lysinuria) and transport disorders  
CC (e.g. myotonic dystrophy, catatonias, peripheral neuropathy). Sequences  
CC ABG99943-ABG60220 represent human DITHP polypeptides of the invention  
XX  
SQ Sequence 239 AA;

Alignment Scores:  
Pred. No.: 1.98 Length: 239  
Score: 89.00 Matches: 38  
Percent Similarity: 31.85% Conservative: 5

Best Local Similarity: 28.15% Mismatches: 48  
Query Match: 10.16% Indels: 44  
DB: 5 Gaps: 7  
US-10-057-510-1 (1-486) x ABG60084 (1-239)  
QY 132 CCAGCTTCGTTGTTCCCT-----CTGGCTTGTGGGGCAGC 167  
DB 18 ProAlaGlyLeuLeuProArgAlaAlaLeuGlnAArgValHisLeuSerCysGlyGly\*\*\* 37  
QY 168 GCTGTTTCATGTGGCAAGTGGAGGCATGACGCTGTGGAGGCGCGTGAAGTGAAG 227  
DB 38 -----GluGluLeuSerGlyGlyArgPheGly-----AlaSerAsnLeuGluProAla 53  
QY 228 GAATGACGAGCCCTGGGAGGAGGCA----- 254  
DB 54 ThrTpvValAspProGlyArgArg\*\*\*Leu\*\*\*SerGlyCysGlyAlaProArgGluLeu 73  
QY 255 -----GAAGCTACGACGGCTGAGGATGAAGATG 284  
DB 74 GlyProProArgSerSerHisSerAlaAspGlyLeuArgGlyArgGlnGlyThr 93  
QY 285 CAGCCCTCGATNGTCCCGAGACTCTCAGGACATTGCC-----AGTCAAGGGTTC 335  
DB 94 GlyValIleGlnThrAlaTrpAlaThrHisTpvLeuProCysHisThrAlaGlyPhe 113  
QY 336 GAGCCACNA---GGNCTTGGCTCATATGCGCATGAGGGGAACCTGCATAGAGCCCCCT 392  
DB 114 SerProSerLeuGlySerGlySerValAlaProSerGlyProLeuLeuAlaSerProLeu 133  
QY 393 CCCTGG-----TTGTGCTCCTGGCCCTCT 416  
DB 134 AlaTyrGlyGlyArgLeuLeuSerSerTpvSerArgProSer 148  
RESULT 10  
AAE35942  
ID AAE35942 standard; protein; 247 AA.  
XX AC AAE35942;  
XX DT 17-JUN-2003 (first entry)  
XX DE Lolium perenne defensin f (LpDEFF) protein.  
XX KW Ryegrass; thionin; TH; thaumatin-like protein; TL; defensin; DEF; ER;  
KW elicitor-responsive protein; disease resistance; plant defence response;  
KW protein storage; pest resistance; genetic marker; gene therapy; antipest;  
KW agricultural; LpDEFF.  
XX OS Lolium perenne.  
XX PH Key Location/Qualifiers  
FT Misc-difference 243 /note= "Encoded by NTC; Xaa corresponds to any amino  
FT acid"  
XX WO200288359-A1.  
XX PD 07-NOV-2002.  
XX PF 01-MAY-2002; 2002WO-AU000539.  
XX PR 02-MAY-2001; 2001AU-00004735.  
XX PA (AGRI-) AGRIC VICTORIA SERVICES PTY LTD.  
XX PA (AGRE-) AGRESEARCH LTD.  
XX PI Spangenberg G, Sawbridge TI, Ong EK, Emmerling M;  
XX WPI: 2003-201227/19.  
DR N-PSDB; AAD54405.  
XX New ryegrass or fescue nucleic acid encoding a thionin, thaumatin-like,  
PT





XX DE Human lung tumour-specific antigen encoded by cDNA #26.  
XX KW Lung tumour protein; lung cancer; cytostatic; vaccine.  
XX OS Homo sapiens.  
XX PN W0200060077-A2.  
XX PD 12-OCT-2000.  
XX PF 30-MAR-2000; 2000WO-US008560.  
XX PR 02-APR-1999; 99US-00285323.  
XX PR 09-AUG-1999; 99US-00370838.  
XX PR 30-DEC-1999; 99US-00476235.  
XX PR 03-MAR-2000; 2000US-00519809.  
XX PA (CORI-) CORIXA CORP.  
XX PI Reed SG, Lodes MJ, Mohamath R, Secrist H;  
XX PF WPI; 2000-638466/61.  
XX DR N-PSDB; AAC79071.  
XX PT Novel lung tumor polypeptides and polynucleotides, useful for detecting,  
XX PT monitoring or treating cancer, especially lung cancer.  
XX PS Claim 1; Page 102-103; 243pp; English.  
XX CC The present sequence is given in a specification relating to compounds  
XX CC for therapy and diagnosis of lung cancer. Polypeptides comprising at  
XX CC least an immunogenic part of a lung tumour protein are disclosed. The  
XX CC polypeptides are useful for inhibiting the development of cancer,  
XX CC especially lung cancer. Samples of T cells expressing the polypeptides  
XX CC may be used to inhibit the development of cancer. The polypeptides are  
XX CC also useful for detecting and monitoring the progression of cancer,  
XX CC especially lung cancer  
XX SQ Sequence 163 AA;  
Alignment Scores:  
Pred. No.: 3.75 Length: 163  
Score: 86.00 Matches: 39  
Percent Similarity: 43.22% Conservative: 12  
Best Local Similarity: 33.05% Mismatches: 50  
Query Match: 10.24% Indels: 17  
Gaps: 3  
US-10-057-510-1 (1-486) x ABA44416 (1-163)  
QY 462 AGGCTTTGCNAAAACACAGGGGNGTANGANCAGTTTCCAGACAGAGCCAGGACC 403  
DB 11 ArgLeuSerLeuArgPro-----SerIleProGluLeuSerGluArgThr 25  
QY 402 ACAACAGGAGGGGGCTCTTATGCAAGTTCCTTCATGCCATATGAGCCAAAGNCCTN 343  
DB 26 SerArgProCysArgAlaSerProAlaSerLeuProSerGlnHisThrSerProAla 45  
QY 342 GTGGTCGAACCCCTTGACTGGGCAATGCC-----TGAGAGTCTGGGGACNATCCA 292  
DB 46 GlnAlaArgValArgAsnLeuAlaGlnSerThrPheProLeuAlaAlaGlnGluThrPro 65  
QY 291 GGGGTGCTATCTTCATCTCAGCCCTGCTAGCCCTTCCTCCCTCCAGGGCTCGTCC 232  
DB 66 GlyArgAlaProAlaHisAlaProLeuSerSerPheValProGlyValGlyArgSer 85  
QY 231 ATTCTTCAGCTCCAGCGCTCTCCACACAGTCCTCATGCTTCCACCTTGCCACATGGA-- 174  
DB 86 -----ProAlaSer-ValGlyIleSerAlaProGlyGlyGlyProSerGlyAl 101  
QY 173 -NACAGCCGTGCCCC-----ACAAGCCAGAGGGGAACAACGAAGCTGS 132

DB 101 aAlaAlaLysIleProLeuGluLeuThrGlnSerArgValGlnLysIleTyr 118  
RESULT 13  
AAE13757  
ID AAE13757 standard; protein; 163 AA.  
XX AC AAE13757;  
XX DT 26-FEB-2002 (first entry)  
XX DE Human lung tumour-specific protein L786-10.  
XX KW Human; lung tumour protein; immunostimulant; cytostatic; gene therapy;  
XX KW antisense-therapy; vaccine; immune response; lung cancer; L786-10.  
XX OS Homo sapiens.  
XX PN W0200172295-A2.  
XX PD 04-OCT-2001.  
XX PF 28-MAR-2001; 2001WO-US009991.  
XX PR 29-MAR-2000; 2000US-00538037.  
XX PR 05-JUN-2000; 2000US-00589937.  
XX PR 18-AUG-2000; 2000US-00640878.  
XX PR 22-SEP-2000; 2000US-0234517P.  
XX PR 01-NOV-2000; 2000US-00704512.  
XX PR 14-DEC-2000; 2000US-00738973.  
XX PA (CORI-) CORIXA CORP.  
XX PI Reed SG, Lodes MJ, Mohamath R, Secrist H, Benson DR, Indirias CY;  
XX PI Henderson RA, Fling SP, Algate PA, Elliot M, Mannion J, Kalos MD;  
XX WPI; 2001-639201/73.  
XX DR N-PSDB; AAD23146.  
XX PT New human lung-specific polynucleotides and polypeptides for the  
XX PT diagnosis and treatment of disease e.g. lung cancer.  
XX PS Example 2; Page 158-159; 378pp; English.  
XX CC The invention relates to isolated lung tumour-specific proteins and their  
XX CC corresponding cDNA molecules. Lung tumour-specific proteins and their  
XX CC antigen-presenting cells are useful for stimulating and/or expanding T  
XX CC cells specific for a tumour protein, and for inhibiting the development  
XX CC of cancer. The invention also relates to a composition useful for  
XX CC stimulating an immune response, and for treating cancer. The lung tumour  
XX CC specific oligonucleotide is useful in gene therapy and for diagnosis,  
XX CC detection and treatment of lung cancer. The present sequence is human  
XX CC lung tumour-specific protein  
XX SQ Sequence 163 AA;  
Alignment Scores:  
Pred. No.: 3.75 Length: 163  
Score: 86.00 Matches: 39  
Percent Similarity: 43.22% Conservative: 12  
Best Local Similarity: 33.05% Mismatches: 50  
Query Match: 10.24% Indels: 17  
Gaps: 5  
US-10-057-510-1 (1-486) x AAE13757 (1-163)  
QY 462 AGGCTTTGCNAAAACACAGGGGNGTANGANCAGTTTCCAGACAGAGCCAGGACC 403  
DB 11 ArgLeuSerLeuArgPro-----SerIleProGluLeuSerGluArgThr 25  
QY 402 ACAACAGGAGGGGGCTCTTATGCAAGTTCCTTCATGCCATATGAGCCAAAGNCCTN 343  
DB 26 SerArgProCysArgAlaSerProAlaSerLeuProSerGlnHisThrSerProAla 45

QY 342 GTGGCTCGAACCTTGACTGGCATGTCC-----TGAGAGTCTGGGACNATCCA 292  
 Db 46 GlnAlaArgValArgAsnLeuAlaGlnSerThrPheProLeuAlaAlaGlnGluThrPro 65  
 QY 291 GGGGCTGCATCTTCATCCTCAGCCCTGGTAGCCTTCTGCGCTCTCCCTCCAGGGGCTGCTCC 232  
 Db 66 GlyArgAlaProAlaHisAlaProLeuSerSerPheValProGlyValGlyArgSer 85  
 QY 231 ATTCTTTCAGCTTCAGCCCTCCACACAGCTCCATCCCTCCATCCCTGCGCATGGA-- 174  
 Db 86 -----ProAlaSer-ValGlyIleSerAlaProGlyGlyProSerGlyAl 101  
 QY 173 -NACAGCCGTGCCCC-----ACAAGCCAGAGGGAACAACAGAGCTGG 132  
 Db 101 aAlaAlaIleProLeuGluLeuThrGlnSerArgValGlnIleTrp 118

RESULT 14  
 ADD66349  
 ID ADD66349 standard; protein; 163 AA.  
 AC ADD66349;  
 DT 15-JAN-2004 (first entry)  
 XX Human lung tumour-specific protein sequence, SEQ ID NO 41.  
 DE expression control; cancer; T cell; tumour; immune; cytostatic; vaccine;  
 KW human; lung tumour-specific.  
 OS Homo sapiens.  
 XX WO200292001-A2.  
 XX 21-NOV-2002.  
 XX 10-MAY-2001; 2002WO-US014975.  
 XX 11-MAY-2001; 2001US-00854133.  
 XX (CORI-) CORIXA CORP.  
 PI Lodes MJ, Wang T, Fan L, Algate PA, McNeill PD;  
 DR WPI; 2003-120592/11.  
 DR N-PSDB; ADD67083.  
 XX  
 PT New polynucleotide and polypeptide, useful for preparing a composition  
 for diagnosing, treating or preventing cancer.  
 XX  
 PS Example 2; SEQ ID NO 41; 494pp; English.  
 XX  
 CC The invention relates to a novel isolated polynucleotide comprising one  
 of 32 47-6080 base pair sequences, given in the specification, or their  
 complements or degenerate variants, at least 20 contiguous residues of a  
 sequence in, or having at least 75 or 90 % identity with the isolated  
 polynucleotide, or that hybridise with the polynucleotide. The invention  
 further comprises: an isolated polypeptide; an expression vector  
 comprising the polynucleotide operably linked to an expression control  
 sequence; a host cell transformed or transfected with the expression  
 vector; an isolated antibody or its antigen-binding fragment that  
 specifically binds to the polypeptide; a method for detecting the  
 presence of a cancer in a patient; a fusion protein comprising the  
 polypeptide; an oligonucleotide that hybridises to the isolated  
 polynucleotide under moderately stringent conditions; a method for  
 stimulating and/or expanding T cells specific for a tumour protein; an  
 isolated T cell population; a composition comprising a first component  
 consisting of carriers and immunostimulants and a second component; a  
 method for stimulating an immune response in a patient; a method for  
 treating cancer in a patient; a method for determining cancer in a  
 patient; a diagnostic kit comprising at least one oligonucleotide or  
 antibody and a detection reagent comprising a reporter group; and a  
 method for inhibiting the development of cancer in a patient. The  
 compositions of the invention have cytostatic activity and can be used to

CC create a vaccine. The isolated polynucleotide is useful for preparing a  
 CC composition for diagnosing, treating or preventing cancer. This sequence  
 CC represents a human lung tumour-specific protein relating to the  
 CC invention.  
 XX  
 SQ Sequence 163 AA;

Alignment Scores:  
 Pred. No.: 3.75 Length: 163  
 Score: 86.00 Matches: 39  
 Percent Similarity: 43.22% Conservative: 12  
 Best Local Similarity: 33.05% Mismatches: 50  
 Query Match: 10.24% Indels: 17  
 DB: 7 Gaps: 5

US-10-057-510-1 (1-486) x ADD66349 (1-163)

QY 462 AGGCTTTGCNAAACACAGGGGNGCTANGANCAGTTTCCAGAACAGAGGGCCAGGACC 403  
 Db 11 ArgLeuSerLeuArgPro-----SerIleProGluLeuSerGluArgThr 25  
 QY 402 ACAACACAGGGGGGCTTCTTATGCAAGTTCCCTTCATGCCATATGAGCCAGNCCTN 343  
 Db 26 SerArgProCysArgAlaSerProAlaSerLeuProSerGlnHisThrSerSerProAla 45  
 QY 342 GTGGCTCGAACCTTGACTGGGCAATGTCC-----TGAGAGTCTGGGACNATCCA 292  
 Db 46 GlnAlaArgValArgAsnLeuAlaGlnSerThrPheProLeuAlaAlaGlnGluThrPro 65  
 QY 291 GGGGCTGCATCTTCATCCTCAGCCCTGGTAGCCTTCTGCGCTCTCCCTCCAGGGGCTGCTCC 232  
 Db 66 GlyArgAlaProAlaHisAlaProLeuSerSerPheValProGlyValGlyArgSer 85  
 QY 231 ATTCTTTCAGCTTCAGCCCTCCACACAGCTCCATCCCTCCATCCCTGCGCATGGA-- 174  
 Db 86 -----ProAlaSer-ValGlyIleSerAlaProGlyGlyProSerGlyAl 101  
 QY 173 -NACAGCCGTGCCCC-----ACAAGCCAGAGGGAACAACAGAGCTGG 132  
 Db 101 aAlaAlaIleProLeuGluLeuThrGlnSerArgValGlnIleTrp 118

RESULT 15  
 ADE87603  
 ID ADE87603 standard; protein; 163 AA.  
 XX  
 AC ADE87603;  
 DT 29-JAN-2004 (first entry)  
 XX  
 DE Human lung tumour antigen polypeptide #10.  
 XX  
 KW Human; lung tumour antigen; cancer; lung cancer; CD4+; CD8+; T cell;  
 XX immune response; immunostimulant; cytostatic.  
 OS Homo sapiens.  
 XX US2003118599-A1.  
 XX 26-JUN-2003.  
 XX 10-MAY-2002; 2002US-00144649.  
 XX 02-APR-1999; 99US-00285323.  
 XX 09-AUG-1999; 99US-00370838.  
 XX 30-DEC-1999; 99US-00476235.  
 XX 03-MAR-2000; 2000US-00518809.  
 XX 29-MAR-2000; 2000US-00538037.  
 XX 05-JUN-2000; 2000US-00588937.  
 XX 18-AUG-2000; 2000US-00640878.  
 XX 20-SEP-2000; 2000US-00667170.  
 XX 01-NOV-2000; 2000US-00704512.  
 XX 14-DEC-2000; 2000US-00738973.  
 XX 11-MAY-2001; 2001US-00854133.

Job time : 59.5 secs

```
XX (CORI-) CORIXA CORP.
XX
XX Algate PA, Lodes MJ, Wang T, Fan L, McNeill PD;
XX
XX WPI; 2003-897103/82.
XX N-PSDB; ADE87588.
XX
XX New polynucleotides encode lung tumor antigens and are useful to
XX stimulate an immune response or detect or treat a cancer in a patient,
XX particularly lung cancer.
XX
XX Example 2; SEQ ID NO 41; 63pp; English.
XX
XX The invention relates to polynucleotides encoding lung tumour antigens.
XX The invention also relates to the polypeptides encoded by the
XX polynucleotides, isolated antibodies or antigen-binding fragments that
XX specifically bind the polypeptides and a method for detecting cancer in a
XX patient, comprising obtaining a biological sample from the patient,
XX contacting the sample with a binding agent that binds a polypeptide of
XX the invention, detecting in the sample an amount of polypeptide that
XX binds to the binding agent, and comparing the amount of polypeptide to a
XX predetermined cut-off value. T cells specific for a tumour protein can be
XX stimulated and/or expanded by contacting the T cells with a polypeptide,
XX polynucleotide or an antigen-presenting cell that expresses a
XX polypeptide. Cancer development can be inhibited by incubating CD4+
XX and/or CD8+ T cells isolated from a patient with a polypeptide,
XX polynucleotide or an antigen-presenting cell that expresses a
XX polypeptide, so that the T cells proliferate. The invention is used to
XX stimulate an immune response or to detect or treat a cancer in a patient,
XX particularly lung cancer. This sequence represents a human lung tumour
XX antigen polypeptide of the invention. Note: the sequence data for this
XX patent did not form part of the printed specification but was obtained in
XX electronic format from USPTO at seqdata.uspto.gov/sequence.html.
XX
XX SQ Sequence 163 AA;

Alignment Scores:
Pred. No.: 3.75 Length: 163
Score: 86.00 Matches: 39
Percent Similarity: 43.22% Conservative: 12
Best Local Similarity: 33.05% Mismatches: 50
Query Match: 10.24% Indels: 17
DB: 7 Gaps: 5

US-10-057-510-1 (1-486) x ADE87603 (1-163)
QY 462 AGGCTTTCNAAACACAGGGGGNCTANGANCAGTTTCCAGAACAGAGGCCAGGACC 403
Db 11 ArgLeuSerLeuArgPro-----SerileProGluLeuSerGluArgThr 25
QY 402 ACAACAGGAGGGGGGCTTCTATGCAAGTCCCTTCATGCCATATGAGCCAGNCCTN 343
Db 26 SerArgProCyArgAlaSerProAlaSerLeuProSerGlnHisThrSerSerProAla 45
QY 342 GTGGCTCGAACCCCTTGACTGGGCAATGTC-----TGAGAGTCTGGGACNATCCA 292
Db 46 GlnAlaArgValArgAsnLeuAlaGlnSerThrPheProLeuAlaAlaGlnGluThrPro 65
QY 291 GGGGCTGATCTTCATCTCAGCCCTCGTACCTTCCTGCTCTCCCTCTCCAGGGCTGCTCC 232
Db 66 GlyArgAlaProAlaHisAlaProLeuSerSerPheValProGlyValGlyGlyArgSer 85
QY 231 ATTCCTTCAGCTCCAGGCTCCTCCACAGCTCCATGCTTCCACCTTCCACATGGA-- 174
Db 86 -----ProAlaSer-ValGlyIleSerAlaProGlyGlyGlyProSerGlyAl 101
QY 173 -NACAGCCGTGCCCC-----ACAAGCCAGAGGGAACAACAGACTGG 132
Db 101 aAlaAlaLysIleProLeuGluLeuThrGlnSerArgValGlnLysIleTrp 118

Search completed: April 6, 2004, 17:55:22
```